1994 DIRECTOR'S REPORT

Division of Computer Research and Technology

National Institutes of Health

Contents

30th Anniversary of DCRT

The Division of Computer Research and Technology is celebrating its 30th Anniversary in 1994. During the past 30 years we have seen enormous changes. The use of computing has increased several thousand-fold; the unit cost has decreased 1,000 fold. The processing speed of our mainframes is more than 1,000 fold what it was in 1964. A desktop PC has more processing speed, memory and disk storage than the mainframe of 30 years ago.

The advent of minicomputers, workstations, PC's, and Macintoshes has totally changed the scene. Now, instead of impacting a few hundred or 1%- 2% of NIH employees, we now have 17,000 users of our central facilities, more than 15,000 PCs or Macintoshes and workstations, 250 Local Area Networks (LANs) involving 8,000 "nodes" on the network.

Further, the capacity of our high performance, highly parallel supercomputers continues to grow, to meet the "Grand challenge" problems of drug design, vaccine design, prediction of protein folding and homology modelling, protein structure determination by x-ray crystallography and multidimensional NMR, multiple sequence alignment, and biomedical image processing, and to meet the "National Challenge" areas such as health care delivery and lifelong learning.

To commemorate the 30th Anniversary, DCRT sponsored a symposium in Masur Auditorium, featuring Dr. Russell Doolittle from the University of California San Diego, and Dr. Martin Karplus of Harvard University. Dr. Doolittle spoke on "The Computer as Biology's Telescope" (looking long ago and far away at molecular evolution). Dr. Karplus described his recent work on "Dynamics of Proteins: from the Native to the Denatured State and Back Again." DCRT's Director, Dr. David Rodbard, provided a brief review of some changes in the computing scene at NIH over the past thirty years. Our DCRT picnic and Awards Ceremony carried over the 30th Anniversary theme.

In view of the transformation of NIH and the radical changes in computing technologies, we at DCRT have transformed ourselves – with a complete reorganization over the past four years. DCRT continues to play several critical roles:

- Providing a vision for the future of computing, networking and information systems at NIH and in biomedical research
- Sponsoring the standardization of computing architectures to promote inter-operability and reduce the costs of procurement and support of computing and networking systems
- Providing shared resources for all of the ICDs and NIH OD, facilitating and supporting their own programs in computing, networking and information systems
- Providing central facilities:
  Mainframes, supercomputers, central support for workstations (electronic software installation, maintenance and support)
- Providing the "enterprise" information and administrative computing systems for support of the "corporate NIH", e.g., the administrative database and its information system and support for both the intramural and extramural programs
- Providing the infrastructure for networking:
  NIHnet-wide area backbone and access to Internet; e-mail services and the central directory; and design and oversight of network cabling of new buildings (e.g., Natcher, Rockledge)
- Providing assistance to users:
  The "4-DCRT" (4dcrt@nih.gov) hotline, for help with PCs, Macs, Network, Mainframe, and Statistics; support of the Scientific Computing Resource Center; and assistance to the Division of Personnel Management (DPM) for design, development and support of the User Resource Centers and the computer/ information systems aspects of the DPM training program
- Organizing a wide range of support groups for users:
  PC Lead Users and Mac Support Coordinators now combined into the Computer Support Coordinators; reorganization in the Distributed Systems Branch; Campus User Research Exchange/Technical Lan Coordinators (CURE/TLCs); ADB Steering Committee; Telecommunications Subcommittee
- Training:
  Offering classes in dozens of subject areas, taken by hundreds of students, and providing a wide range of self-instructional materials used by hundreds more
- Providing assistance with evaluation, procurement, installation of hardware, software and support services including pioneering of site licensing and bulk purchases

- Providing leadership in moving NIH to the forefront of biomedical research information systems, e.g., with MOSAIC, World Wide Web and GOPHER with support for molecular biology databases and sequence analysis, and the development of the NIH Molecular Modeling Home Page" and "Molecules 'R Us"
- Conducting research at the interface of computer science and biomedicine and in related fields of structural biology, image processing, biomathematics, biophysics and biostatistics
- Collaboration and consulting with all of the ICDs.

The use of the Internet has hit numerous magazine covers from *Science* to *Time* to the *Communications of the ACM*. There is now an exponential growth of number of sites, amount of traffic, and the volume and quality of the information. DCRT has played a major role in stimulating and fostering this at NIH: setting up prototypes, introducing NIH employees to the possibilities, training, instructional booklets, support and liaison.

Director's Preface: FY'94 Highlights at DCRT

1994 was another exciting and enormously productive year for DCRT. Some of the highlights include:

- *Architectural Management*: The proliferation of computing platforms, network operating systems and applications programs gives rise to a combinatorial explosion, aggravated by the need to support multiple versions of each product. We seek the elusive goal of providing "seamless interoperability" for these heterogeneous systems, e.g., for client-server applications. Unless this architectural chaos is brought under control, it will become progressively more difficult and ultimately impossible to support computer systems at NIH! To begin the difficult process of developing a consensus architecture for as many systems as possible, DCRT initiated an "Architectural Management Retreat" for ourselves in October 1993. This was so successful that DCRT organized a similar retreat for representatives of all of the ICDs in February 1994. The final report, delivered in September 1994 is ready to be presented to senior NIH management.
- *Technical Assistance and Support Center (TASC)*: A new approach to providing service to our DCRT customers was initiated February 1, 1994, with the inauguration of the TASC. TASC provides "one-stop shopping": one can simply call 4-DCRT or e-mail to 4DCRT@nih.gov to obtain consultation services for mainframe, networking, PCs, database, statistics, the TARGET center for computer technologies to assist employees with a wide range of disabilities, and other DCRT services. TASC has proven to be enormously successful, and this central service will be expanded to the extent resources permit. TASC is operated by the *Customer Service Branch*, which also organizes the highly successful *DCRT Training Program* and *Scientific Seminar Series*.
- Our *Network Systems Branch* has taken a leadership role in providing the copper and fiberoptic cabling for the NIH backbone, and for vertical and horizontal wiring for new facilities, e.g., the Natcher Building and the Rockledge Building. The new NIH Central Directory service for e-mail and phone addresses has been an enormous success.
- The *Computational Molecular Biology Section* and the *Scientific Computing Section* of CFB have put NIH and DHHS on the map in terms of providing important new services over the Internet, using *GOPHER* and *MOSAIC*. For example the *NIH Home Page for Molecular Modeling* was featured on the cover of Science (1), and provides access to "Molecules 'R Us" - an encyclopedic review of available molecular modeling software.
- The *Computing Facilities Branch* has made substantial progress in modernizing its facilities and services, including a major overhaul of its terabyte of disk storage, moving to the new technology of *Redundant Arrays of Inexpensive Disks (RAID)* for Direct Access Storage Devices (DASD) and completion of a major study leading to *automation* of console control, system managed storage, and printer operations. The Byzantine process for *procurement* of major computer equipment and services is moving along well with the assistance of GSA's FEDSIM and FEDCAC: a *concept of operations* and a detailed *requirements analysis* have been completed, including interviews or workshops involving more than 300 NIH employees representing all constituencies: the scientific, administrative, intramural and extramural programs. The Convex Scientific Computing system has been substantially augmented by the addition of a Silicon Graphics Challenger, providing considerable cost savings.
- The *Computational Bioscience and Engineering Laboratory* has continued to be an award winning, world-class leader in the application of *highly parallel high performance computing* to important biomedical problems, ranging from image processing (both laboratory and clinical), genetic database searching, protein structure prediction algorithms to radiotherapy planning and optimization. A summary of much of this work appeared in the *Computers '94 issue of Science.* (2). Together with collaborators in NIAMS, computer enhanced image processing has been used to make some remarkably sensitive and subtle measurements regarding the exact location of various proteins involved in viral assembly (3).
- Science at the interface of computing, mathematics, physics, biology and chemistry is conducted by the *LSB* and *PSL*. Notable accomplishments include: measurement of the exchange of chloride ions (along with water) during the oxygenation of hemoglobin (4), use of naturally occurring ion-channels as a "Coulter counter" to measure the

transport of particles (5). The MGSS has been a pioneer in the use of *clusters of high-performance workstations* on a high speed network. The *CHARMM* molecular dynamics package, involving 150,000 lines of code, has been successfully ported to the Intel i860 Highly parallel supercomputer, and to a large cluster of high performance workstations. Significant progress has been made in understanding the possible *mechanisms of action of HIV protease* (6), *the design of HIV protease inhibitors*, and in understanding the *structural correlates with syncytium induction by virulent strains of HIV* (7). Members of the Division have authored or edited *three major monographs* in addition to a prolific series of articles, (8-10).

- DCRT's newly constituted *Information Systems Branch* is developing new user-friendly screens for the ADB and *ADB Information System* (ADBIS). Further, DCRT has been contributing leadership for *"Business Process Reengineering"* to streamline and simplify administrative processes for all of NIH. Rather than automating or "computerizing" outmoded inefficient processes, the goal is to abolish as many obsolete forms and procedures as possible, and then automate only the essential ones in the most efficient possible manner.
- The *Distributed Systems Branch* has provided yeoman's service while undergoing an extensive *reorganization*: massively outnumbered by the 15,000 or more users of distributed computing, they are positioning themselves to provide those services that should be performed centrally from a cost-effectiveness standpoint, and to provide support to the computer and networking experts on the local (ICD) levels. DSB has opened an *Image Processing Technology Center* for use by scientists throughout the campus.
- *Equal Employment Opportunity*: DCRT has initiated a new program, in collaboration with the NIH OEO, to provide all NIH employees access to the TARGET Center of the Department of Agriculture. This Center provides computer and other technologies to assist persons with disabilities. Examples of the facilities are: voice activated response, talking terminals, assists for paraplegic or quadriplegic individuals and for those with repetitive stress disorders. The *Career Enhancement Program* of our *Human Resources Management Office* and the *Computer Facilities Branch*, is a highly successful program to provide expanded opportunities for professional advancement to persons in positions such as computer operators. We are planning to expand this program.

DCRT is doing everything it can to promote the appropriate, optimally effective and cost-effective use of computing, networking, and information systems by all sectors of NIH. With the exponential growth of new services, e.g., GOPHER and MOSAIC and other information systems, and the advent of inexpensive desk-to-desk videoconferencing, the potential for computing in biomedical research has once again been expanded by one to two orders of magnitude.

DCRT has had an extraordinarily productive and successful year despite the presence of significant resource constraints. We are continually seeking new collaborative and cooperative projects with all of the ICDs, and looking for opportunities to be of even better service to the entire NIH community.

David Rodbard, M.D.
Director, DCRT

References

1. Science, August 12, 1994;Vol. 265. Computers '94 Networks and Modeling.
2. Martino RL, Johnson CA, Suh EB, Yap TK. Parallel computing in biomedical research. Science 1994;265:902-908.
3. Booy FP, Trus BL, Newcomb WW, Brown JC, Conway JF, Steven AC. Finding a needle in a haystack: detection of a small protein (the 12 kDaVP26) in a large complex (the 200-Mda capsid of herpes simplex virus). Proc Natl Acad Sci USA 1994;91:5652-5656.
4. Colombo MF, Rau DC, Parsegian VA. Re-evaluation of chloride's regulation of hemoglobin oxygen uptake: the neglected contribution of protein hydration. Proc Natl Acad Sci USA 1994;91:10517-10520.
5. Bezrukov SM, Vodyanoy I, Parsegian, VA. Counting polymers moving through a single ion channel. Nature 1994;370:279-281.
6. Chatfield DC, Brooks BR. HIV-1 proteinase cleavage mechanism elucidated with molecular dynamics simulation. (submitted).
7. Bhattacharyya D, Brooks BR, Callahan L. Positioning of positively charged residues in the

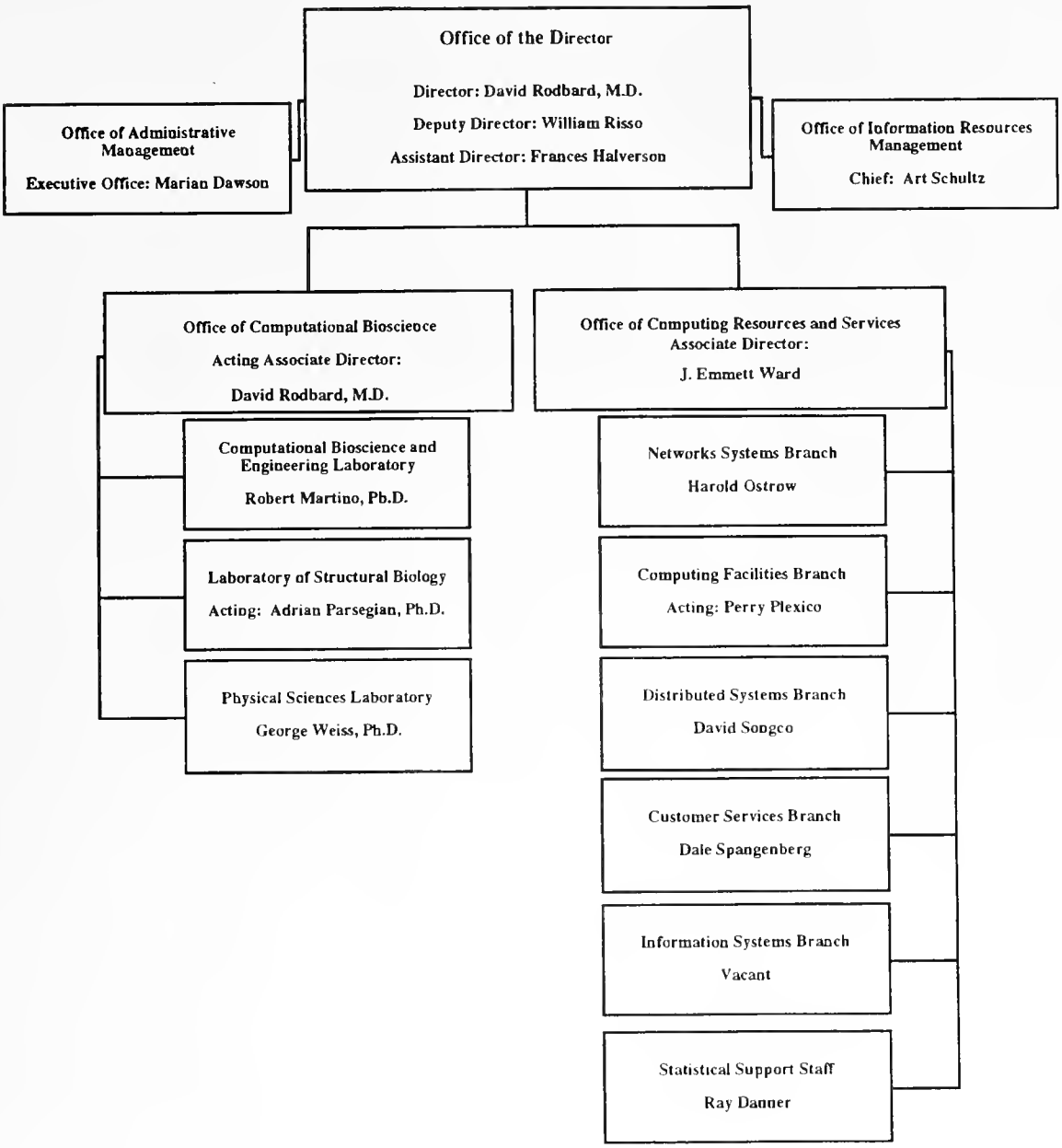
V3 loop correlates with HIV-1 syncytium-inducing phenotype. (submitted).

8. Shmueli U, Weiss GH. Introduction to Crystallographic Statistics. Oxford: Oxford Univ Press, 1994 (in press).

9. Weiss GH. Aspects and applications of the Random Walk. Amsterdam: North-Holland Press, 1994;361.

10. Malley J. Statistical applications of Jordan algebras. Research monograph in *Lecture Notes in Statistics Series*, Vol 91. New York: Springer-Verlag, 1994.

DCRT ORGANIZATIONAL STRUCTURE



Office of Computational Biosiences

Computational

Bioscience and

Engineering

Laboratory

Computational Bioscience and Engineering Laboratory

Robert L. Martino, Ph.D., Chief

The Computational Bioscience and Engineering Laboratory (CBEL) is devoted to the exploitation of high performance computer systems in biomedical applications, including image processing, structural biology, computational chemistry, medical imaging, scientific visualization, signal processing, genetic database searching, genetic linkage analysis, and advanced statistical methods. Members of CBEL strive to identify and solve those computational problems in biomedicine that can benefit from high performance hardware, modern software engineering principles, and new and efficient algorithms. The Laboratory also provides high performance parallel computer and image processing systems for the NIH scientific staff and collaborators.

CBEL provides leadership in the research, development, and biomedical application of massively parallel computers in a networked environment. It collaborates with research investigators in modeling of complex systems, analyzing and interpreting data, signals and images, and assisting with computationally intensive tasks in electron and light microscopy, x-ray crystallography and NMR spectroscopy for protein structure determination, molecular dynamics and quantum chemistry in the design of pharmaceuticals, medical imaging to study brain function, and in radiation treatment planning for the treatment of cancer. CBEL conducts continuing research to expand the use of high performance computing in biomedical areas. It develops research systems into progressively more accessible and user-friendly systems which ultimately become routine computer utilities.

CBEL's High Performance Biomedical Computing project is an important part of the national, multiagency High Performance Computing and Communications (HPCC) Program. The Information Infrastructure Technology and Applications (IITA) component was added as an important new initiative of the HPCC Program during FY94. In response to this initiative, CBEL and the Distributed Systems Branch of DCRT, in collaboration with the NIH Clinical Center and the National Cancer Institute, are implementing a prototype high-speed network, capable of supporting multimedia communication for medical research and education, on the NIH campus. This network environment will support high-performance radiation therapy planning, which is a collaborative effort between DCRT and the NCI Radiation Oncology Branch (ROB). Connected to this network, DCRT's Intel® iPSC/860 Supercomputer will be utilized to apply the power of parallel computing methods to the computationally intensive calculations required for radiation therapy planning. Special-purpose medical workstations will be deployed as nodes to the prototype network in other selected domains on the NIH campus. DCRT will use the Asynchronous Transfer Mode (ATM) technology as the basis for the prototype high-speed network. CBEL also participated in the ARPA High Performance Computing and Communication Symposium and served on the organizing committee for the Pasadena II Workshop on System Software and Tools for High Performance Computing Environments.

During FY94, CBEL worked closely with DCRT's Office of Computing Resources and Services (OCRS) to plan for the acquisition and deployment of future computing systems and services. CBEL served on the DCRT Architecture Management Staff, which provides leadership, guidance, and coordination for the overall computing and networking environment that DCRT offers and supports for the NIH user community. CBEL also worked with OCRS employees on the requirements analysis for the Computing Equipment Resources and Technology Acquisition for NIH (CERTAN), concentrating on the area of scientific computing.

CBEL executes its work through an Office of the Chief and two operating sections:

- The *Office of the Chief* provides overall CBEL management and planning, including laboratory, administrative and financial functions. It coordinates the establishment of new laboratory activities and the work of the sections to encourage and ensure appropriate cooperation and integration of effort. It also coordinates CBEL work with other parts of DCRT and the NIH ICDs, as well as other government agencies and research institutions.
- During the past year, Calvin Johnson was named the new Chief of the *High Performance Computing Section* (HPCS). This section develops high performance computer systems for the solution of biomedical laboratory and clinical research problems. It provides parallel algorithm expertise to solve computationally intensive problems in biomedicine. HPCS deploys modern, non-traditional, computer architectures in a distributed computing environment, and provides a high performance parallel computer facility for the NIH scientific staff.

- Benes Trus, who holds a joint appointment with the Laboratory of Structural Biology Research, NIAMS, is the Chief of the *Image Processing Research Section* (IPRS). IPRS creates and adapts algorithms and computational techniques for various biomedical imaging modalities including electron microscopy, light microscopy, Positron Emission Tomography (PET), Single Photon Emission Computer Tomography (SPECT), and Magnetic Resonance Imaging (MRI). It performs research in structural biology and biochemistry, using state-of-the-art image processing methods. IPRS also provides an image processing facility for CBEL and other collaborating laboratories.

Research Projects

High Performance Biomedical Computing

R.L. Martino, Ph.D. and C.A. Johnson with J.C. Pfeifer, E.B. Suh, B.L. Trus, Ph.D., N.I. Weisenfeld, T.K. Yap, C.J. Lanczycki (DCRT/CBEL); J.I. Powell, K.M. Kempner (DCRT/DSB); B.R. Brooks, Ph.D. (DCRT/MGSS); A.C. Steven, Ph.D., J.C. Conway, Ph.D., F.P. Booy, Ph.D., J.R. Caston, Ph.D. (NIAMS/LSBR); J. van de Geijn, Ph.D., Huchen Xie, Ph.D. (NCI/ROB); M.E. Daube-Witherspoon, Ph.D., R.E. Carson, Ph.D., Y. Yan (CC/PET); J.V. Haxby, Ph.D., J. Maisog, M.D. (NIMH/LPP); J. Giedd, M.D., J. Rajapakse, Ph.D. (NIMH/CPB); R. McIntosh, Ph.D. (NIA/LN); S. Ghosh, M.D. (NCHGR); E.S. Gershon, M.D., Ph.D., L.R. Goldin, Ph.D. (NIMH/CNB); B. Lee, Ph.D. (NCI/LMB); S. Erwin, T. Mattson, Ph.D. (Intel Supercomputer Systems Division); J. Saltz, M.D., Ph.D., R. Das, Ph.D. (University of Maryland); A. Toga, Ph.D., R. Chann (UCLA School of Medicine); T.S. Baker (Purdue University); O. Frieder (George Mason University); J. Ott, Ph.D. (Columbia University); B. Venkataraghavan, Ph.D. (American Cyanamid Company); R. Parker, Ph.D., J.C. Toole (ARPA/CSTO)

The goals of the high performance biomedical computing program are to identify and solve those computational problems in biomedicine that can benefit from high performance hardware, modern software engineering principles, and efficient algorithms. This effort includes providing high performance parallel computer systems for the NIH staff and developing parallel algorithms for biomedical applications.

Using high performance parallel computers, biomedical scientists can greatly reduce the time it takes to complete computationally intensive tasks and take new approaches in processing their data. This may allow the inclusion of more data in a calculation, the determination of a more accurate result, a reduction in the time needed to complete a long computation, or the implementation of a new algorithm or more realistic model. With proper computer network connections and interactive user interface, parallel computing is readily available to a biomedical researcher in the laboratory or clinic at the investigator's computer workstation.

In addressing these computational challenges, CBEL is developing algorithms for a number of biomedical applications that can benefit from computational speedup, including image processing of electron micrographs, radiation treatment planning, medical imaging, protein and nucleic acid sequence analysis, human genetic linkage analysis, protein folding prediction, nuclear magnetic resonance spectroscopy, x-ray crystallography, quantum chemical methods, and molecular dynamics simulations. The ultimate goal is to have high performance parallel computing facilitate the science that is done at NIH. While developing these computationally demanding applications, CBEL is investigating the following high performance computing issues: partitioning a problem into many parts that can be independently executed on different processors; designing algorithms so that delays of inter-processor communication can be kept to a small fraction of the computation time; designing the parts so that the computing load can be distributed evenly over the available processors or dynamically balanced; designing algorithms so that the number of processors is a parameter and the algorithms can be configured dynamically for the available machine; developing tools and environments for producing portable parallel programs; monitoring system performance; and proving that a parallel algorithm on a given machine meets its specifications.

As part of its high performance computing activity, CBEL operates the DCRT Highly Parallel Computer System. This Intel® Supercomputer Systems Division iPSC/860, obtained in collaboration with the DARPA Touchstone program, is a multiple instruction stream, multiple data stream (MIMD) distributed memory system. The system has 128 processor nodes with 16 megabytes of memory per node. A high-speed data pathway connects all the nodes of the system. Using a hypercube network topology, this hardware message routing facility connects each node as if it had a dedicated channel to all other nodes. Another important part of the system is the Intel® Concurrent I/O File System, which provides a 10 gigabyte fast access mass storage facility for balancing disk input/output with the computational power of the processors. This consists of many small disks connected to I/O nodes that communicate with the processor nodes through the hypercube interconnect. Over the next three years, CBEL will be adding a next-generation high

performance parallel computer, capable of providing up to 100 gigaFLOPS of computing performance.

The President's Office of Science and Technology Policy (OSTP) with the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) support a multiagency High Performance Computing and Communications (HPCC) Program, with the goal of accelerating the development of future generations of high performance computers and networks and of using these resources throughout the American economy. Within the Department of Health and Human Services (DHHS), the focal point for the HPCC program is the National Institutes of Health. The CBEL high performance biomedical computing program is an important part of this national initiative.

Collaborative High Performance Biomedical Computing Projects

Image Processing of Electron Micrographs

High resolution cryoelectron microscopy, in combination with 3-D computer image reconstruction, allows the structure of large icosahedral virus capsids to be studied. The 3-D reconstruction of such viruses begins with one or more electron micrographs, in which each particle on the micrograph is a 2-D projection of a virus specimen. The relative orientation of each particle, denoted by a polar and azimuthal angle pair, defines the angle or view of the corresponding projection. Neglecting the effect of noise and particle imperfections, the virus particle projections in the 2-D micrograph can be considered identical in 3-D except for their orientation. The reconstruction process takes advantage of the icosahedral symmetry of these viruses and the Fourier slice theorem, which states that the orientation of the particle plane in 3-D Fourier space is identical to that of the projection plane. In Fourier space, each 2-D projection is equivalent to 59 other symmetry-related views due to the icosahedral symmetry of the capsid, and these equivalent views all intersect at the Fourier origin. For each projected particle, the reconstruction algorithm interpolates that projection's Fourier transform, in a plane determined by its orientation parameters, onto a 3-D grid in Fourier space.

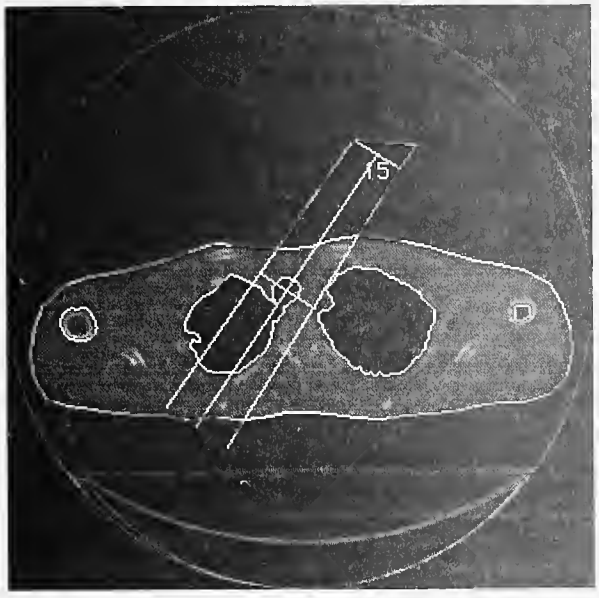
Prior to performing the reconstruction, the orientation parameters of each particle need to be determined with reasonable accuracy. Two separate processes are employed to find initial estimates and then refine the particle orientations, respectively. Reconstruction and the orientation determination are limited by the power of computation ("compute bound") and require high performance computer technology for reconstructions that include many particles. The orientation determination processes were conducted in parallel in previous years.

The major FY94 development has been the design of a parallelization strategy for the reconstruction after the orientation of the particles are determined. This scheme is outlined as follows. 1) Distribute the available set of micrographs across the processors of the parallel computer, combining the data from all processors according to the Fourier projection theorem. 2) Re-distribute the data against the reciprocal Z-coordinate across the processors to facilitate the inversion of a large number of variably sized matrices and perform a Fourier-Bessel transform in the reciprocal radial coordinate. 3) Re-distribute the data against the real-space radial coordinate and perform the remaining two-dimensional inverse transform that yields the full three-dimensional map in cylindrical coordinates. 4) Interpolate the cylindrical coordinate map back to Cartesian coordinates.

The limited amount of usable memory available on the iPSC/860 presents a potential resolution-limiting factor. The memory requirement for reconstruction is roughly proportional to the cube of the desired Fourier cutoff radius, but is independent of the number of particles in the reconstruction. By contrast, in the orientation refinement process known as Emicograd, the primary limiting factor is the number of particles allowed in a refinement. The memory requirements of Emicograd dictate that the particles be distributed across the processors in the parallel implementation. This technique dramatically increases the number of particles that may be included in a reconstruction. Our discovery (in collaboration with NIAMS and the University of Virginia, Charlottesville) of the location of individual VP26 proteins on the hexomers of the human herpes virus capsid was due, in part, to an improvement in the signal-to-noise ratio of recent reconstructions as a result of refining larger numbers of particles. In FY95 we plan to complete the parallel implementation of the reconstruction process. We hope to begin the parallel adaptation of a new code from Purdue University, which refines the orientation of particles against an idealized model.

Radiation Treatment Planning

In radiation treatment planning, a radiation oncologist tries to determine the optimum placement, blocking and intensity of beams such that the body volume to be irradiated receives the maximum dose, while minimizing damage to surrounding tissue. Computational methods can be used to greatly improve the success of this task. A series of images, such as



A source image slice with a beam placed and some contours drawn. The contours denote regions of different density and are subsequently used in the radiation dose calculation in place of the source image. The beam specifies the path of the central ray, width, placement and the presence of the blocking wedge.

from a CAT scan, are read into the computer, and then different volumes are identified. These would typically include bone, lung, internal organs, spinal cord, and the tumor. A beam placement plan is then specified, and the simulation is performed by the computer. This simulation results in a series of 2-D density maps showing the relative radiation absorption, together with isodose contour maps. The radiation oncologist can use the contour maps to compare various beam plans, and thus determine which is the most effective.

CBEL is using the "Projective Beam Model" developed by Jan van de Geijn of NCI (ROB) and others. In this model, beam profiles measured for some reference condition are distorted though a series of quasi-optical functions to determine the percentage of maximum dose at a given point. These functions account for differences from the reference condition in terms of field size and shape, depth, source to surface distance, scattering and changes in density. This model leads to a program with many calls to compute intensive transcendental functions. Currently, the simulation requires that all the beams lie parallel to the plane of the source images. This is a serious limitation, and better treatments plans could be obtained if oblique beams were simulated. Van de Geijn's group is currently developing a full 3-D model to provide this improvement. As it will be some time before their new method is available, CBEL is developing a method to reslice the contour set in an oblique plane, so that the current model could then be used to compute the partial dose due to that beam. That partial dose would be resampled back into the original orientation and added to the total dose.

In FY94, significant progress has been made on implementing a method for handling the oblique beam problem. Also, the core dose calculation was extracted and ported to the Intel® iPSC/860 parallel computer. The source code used for this porting is MacTPS, a Macintosh® based radiation treatment planning system developed by van de Geijn and Huchen Xie of NCI. The Macintosh® provides a good graphical interface for preparing the images and for specifying the beam plan, but the simulation of the treatment takes unacceptably long. This is the first time CBEL has used a Macintosh® computer as a front-end for a program running on the Intel® iPSC/860. There are two difficulties with this: (1) communications between the two machines; (2) allocating a cube and starting the node program. The first is handled on the Macintosh® by a library that implements TCP/IP Internet sockets. The second is handled by a server daemon running on a SUN® workstation.

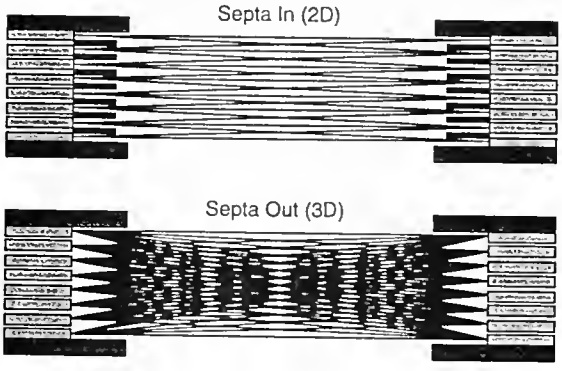
Radiation treatment planning represents a major commitment by DCRT to address the requirements of high performance computation and communication in health care delivery. In the coming year, CBEL will port additional features of MacTPS to the Intel® iPSC/860 parallel computer and investigate automated beam plan optimization. CBEL is also collaborating with the Distributed Systems Branch (DSB) and the Network Systems Branch (NSB) on the Radiation Consultation Workstation project, which will include parallelized treatment planning software. This project will use ATM switching technology over a high speed network to connect multiple researchers at multimedia workstations, medical image servers and the Intel® iPSC/860, enabling researchers at different localities on campus to interact in real–time through both sound and video, as well as giving them a very fast connection to the parallel computer. In addition, we are collaborating with J. Saltz and others at the University of Maryland, who will be investigating advanced computational topics in radiation treatment planning.

Reconstruction of Positron Emission Tomography Images

A Positron Emission Tomography (PET) image is formed through a computational reconstruction process. The quality of the resulting image and the computation time required to produce it depend on the chosen reconstruction algorithm. Traditionally, Fourier methods (e.g., the filtered backprojection method, which

is fast but can lead to artifacts) have been used. The expectation maximization (EM) method is an iterative solution based on a maximum likelihood criterion. The EM method is known in principle to yield more accurate reconstructions or equivalent reconstructions with lower patient dose than the filtered backprojection. The heavy computational demand of the EM method has, however, limited its use.

With the availability of high performance computing, NIH scientists can now consider applying the EM algorithm to the problem of truly 3-D reconstruction. The new generation PET scanners allow for the retraction of the tungsten septa shields, which prevent coincident events from being detected outside the axial plane of emission. Retracting the septa increases the angle over which coincidence events can be accepted, and consequently improves detector sensitivity and count rate. However, the amount of detected scatter and random events also increases with wider acceptance angles. Another drawback to retracting the septa is that the size of the reconstruction grows geometrically when using the EM algorithm. In a 3-D EM reconstruction, using typical scanner geometries, the number of projections (ray coincidence events) grows by an order of magnitude, and the size of the probability matrix, which is used throughout the reconstruction, can grow by four orders of magnitude or more. The current investigation seeks, therefore, to determine whether retracted septa scanners can lead to improved image reconstruction quality.



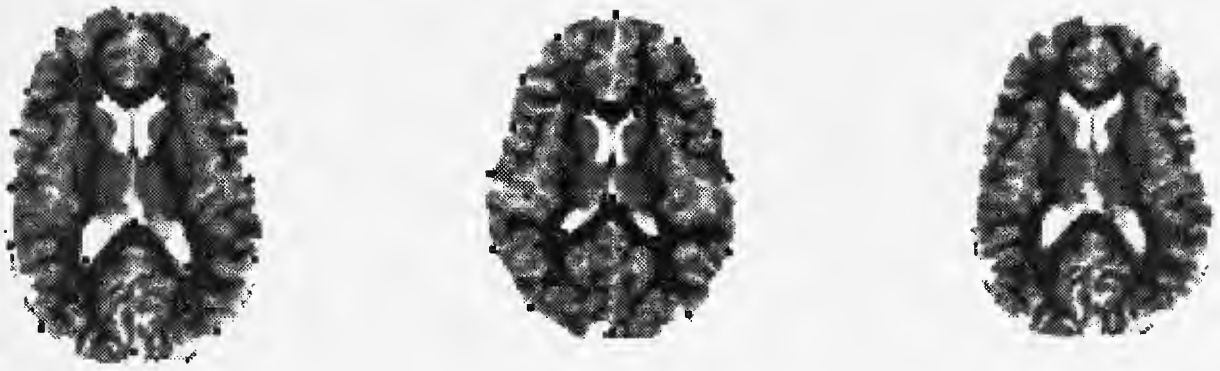
This figure illustrates the effect of retracting the tungsten septa shields on the number of coincidence events to be collected and subsequently processed for reconstruction. Shown is the axial cross section of an 8-ring PET scanner, along with the lines of response for septa-in and septa-out acquisition. Each line represents one sinogram, viewed from the cross-sectional angle. When the septa are in place, only in-plane events and events between neighboring planes are collected, requiring 2x8-1=15 sinograms for an 8 ring detector. With the septa retracted, all possible coincidence events are collected, resulting in 8²=64 sinograms for an 8 ring detector. This type of geometric growth in observed data becomes especially severe in modern PET systems with many detector rings, and 100 million or more chords of coincidence pairs (projections) may be involved in the reconstruction. The geometric growth in observed projection data causes the probability matrix to become larger than can be stored, and prohibitive to compute. Fortunately, redundancy in the PET geometry and a few reasonable assumptions allow for significant reductions in the matrix size. Adapted from Cherry SR, Dahlbon M and Hoffman EJ; J Assist Tomogr 1991;vol 15.

During FY94, we implemented the EM reconstruction algorithm for volume acquisition from current generation retracted-septa PET scanners. Although the system was designed for a GE Advance scanner, the software is easily adaptable to other 3-D scanners. The reconstruction software was written for the Intel® iPSC/860 parallel computer with 128 compute nodes. Running on 32 nodes, the algorithm requires approximately 55 minutes per iteration to reconstruct a brain image. No projection data compression schemes or other approximations were used in the implementation. Extensive use of the probability matrix symmetries (including the 8-fold in-plane symmetries, 2-fold axial symmetries, and axial parallel line redundancies) reduces the storage cost by a factor of 188. Symmetry operators copy and index the probability matrix section to the form required for the particular symmetry. The use of asynchronous reads, lookup tables, and slice-minor image indexing improves computational performance.

Without the extensive use of symmetries and efficient storage of probability matrix sections, the storage requirements of a retracted-septa implementation would be prohibitive. The feasibility of performing such reconstructions in clinical production on the current-generation parallel computer is marginal, given the heavy computational requirements. For larger reconstructions of body-sized images, we look to improvements in parallel computer technology and algorithm design. These larger reconstructions will require more memory, more and faster disk storage, faster processors, and a significant increase in the number of processors. We shall begin preparations for implementing a body-sized reconstruction in the coming year. Corrections for attenuation, random events and scatter will be added to the 3-D code in order to reconstruct from real data. The results of studies on 2-D (septa-in) versus 3-D (septa-out) convergence rates have been inconclusive so far. In FY95, we shall perform more extensive convergence studies and investigate methods for convergence acceleration.

Functional Neurological Image Analysis

CBEL is investigating the use of high performance computing technology and image processing techniques to solve problems in the analysis of medical images that depict neurological function. This work was begun



Single slices of MRI scans of two normal children of different ages. The leftmost scan is warped to have the form of the center scan, using tie-points identified by black squares. The warped image is shown on the right. Data courtesy of J. Giedd and J. Rajapakse (NIMH).

with researchers from NINDS and NIA, and has continued with researchers from NIMH and the UCLA School of Medicine. Our current efforts are focused on a software solution to the problem of "warping" brain images to a standard stereotactic atlas.

In order to implement software systems that analyze brain image data acquired from different subjects, it is necessary to somehow compensate for the variations in brain size and shape between individuals, as people age and due to disease. The proposed solution to this problem is to perform a series of linear and non-linear transformations on each of the acquired images in order to render them into a standard stereotactic space. The end result of such transformations would be that a given image coordinate would reference the same anatomical structure, or functional information derived from the same anatomical structure, in each image. The process of elucidating and applying the proper linear and non-linear transformations is called "warping."

CBEL has been collaborating with researchers from NIMH and the UCLA School of Medicine to implement and refine UCLA's novel warping method on DCRT's 128-node Intel® iPSC/860 massively parallel supercomputer. In FY94, CBEL successfully ported the UCLA warping code to its iPSC/860, and developed a novel data distribution and load balancing algorithm, which was used in the parallel code. Significant speed increases were achieved by applying parallel computing technology.

In FY94, CBEL also began its own research into possible warping algorithms. Such research has been progressing at NIH and elsewhere for at least 10 years; an optimal automated solution, however, has yet to be found. This is considered to be an important hurdle for automated processing of brain images. The use of parallel computing technology for this problem allows the exploration of computationally intensive algorithms, which may someday provide a definitive solution to this problem.

In FY94, CBEL continued refining and maintaining CBEL's high performance implementation of portions of Hammersmith Hospital's Statistical Parametric Mapping (SPM) codes. The plastic normalization portion of CBEL's SPM saw some added features, as requested by some of its current users. The code, which calculates and applies a warping to PET scans, has been used by twenty scientists at NIH, and has been distributed to two non-NIH research centers.

In FY94, CBEL also completed development of software to perform a Principal Component Analysis (PCA) on functional brain images. PCA allows researchers to uncover experimental effects present in image data, but that are buried under other effects not correlated with the experiment. CBEL implemented a computationally efficient PCA algorithm that drastically reduces the number of calculations necessary, while providing the same numerical solutions. CBEL's C-language implementation of the SPM plastic

normalization and CBEL's PCA implementation both provide significant speed increases, allowing researchers to process more studies in substantially less time than they could with conventional codes. CBEL also investigated methods, including learning vector quantitation, for segmenting MRI brain images into regions of gray matter, white matter, and cerebrospinal fluid. The resulting segmented maps can be used as structural templates for functional neuroimaging studies.

In FY95, CBEL plans to continue its exploration of warping algorithms and other tools for functional neurological image analysis. CBEL will also continue to support the use of its software by consulting with end users and, when necessary, providing solutions for any problems that may arise.

Protein and Nucleic Acid Sequence Analysis

Advances in biotechnology have resulted in the generation of a massive volume of biological sequence data, such as that found in the GenBank and SWISS-PROT databases. These databases contain a large number of genetic and protein sequences and their associated biological and bibliographical information. For example, the current release of GenBank (rel. 83.0) contains a total of 182,753 sequences consisting of 191,393,939 bases. Retrieving homologous sequences from such databases is a critical tool for the biomedical research community. When we discover new sequences, we are eager to search these databases for sequences that are similar or related to these discoveries. The biological information associated with the similar sequences found in the database may provide important clues to the structure and function of the newly discovered sequences.

To find homologous sequences, the query sequence must be compared with all the sequences in the database. To obtain high sensitivity, most researchers use the computationally intensive dynamic programming algorithm developed by Smith and Waterman for comparing two sequences. As the size of the databases continue to grow exponentially, it becomes impractical to use the full dynamic programming algorithm on a conventional sequential computer system. To reduce search time, we exploited the power of our massively parallel computer system. In parallelizing this application, we developed a domain decomposition technique. In this method, the original sequences in the database are placed into one of the p buckets (smaller databases), where p is the largest number of possible processors, so that the difference between the sum of the sequence lengths in the smallest and the largest buckets is minimized.

The algorithm for placing the database sequences in these buckets is as follows: first, the sequences are sorted in decreasing length order; then, starting from the longest one, each sequence is placed into the bucket that has the current smallest sum of sequence lengths; in the case of a tie, the smallest numbered bucket is selected; once the database has been decomposed, each processor can search its own bucket independently without communicating with other processors; if only p/n processors are used, each processor can search n buckets.

To evaluate our method, we compared it with the state-of-the-art method presented by a group of researchers at Yale University. Unlike our method, they adapted a dynamic load distribution technique instead of a static one. Their method uses one processor as the master to distribute the sequences to the workers (the remaining processors) for comparison and to collect the similarity scores back from them. The main job of the master is to keep the workers busy as long as there is work to be done. That is, the master assigns a new sequence, as soon as possible, to the worker that completes its assigned comparison. This form of dynamic load distribution continues until all the sequences have been compared. The sequences in the database were sorted to ensure that the last sequence compared is the shortest one.

Our bucket method demonstrated improvement over the master-worker method for a range of query sequence lengths from 50 to 500 bases and a range of processors from 1 to 128. Considering the shortest sequence (50 bases) as an example, it took 3.27 hours to search GenBank (rel. 80.0) for the most homologous sequence on a single processor. The same query took 2.33 minutes using our bucket method and 5.96 minutes using the master-worker method on 128 processors. Therefore, our method achieved more than 155% improvement over the master-worker method for this query.

Another important tool in genetic sequence analysis is multiple sequence alignment. In practice, a number of N similar sequences are often retrieved, where N is usually between 5 and 50. These N sequences may belong to the same family. In the past year, we developed a prototype parallel computational approach for aligning these N sequences to examine their similarities in detail . We want to know which parts of the sequences all the family members have in common and how the family members differ in other parts of the sequences. A multiple sequence alignment is a configuration that is obtained from stacking up sequences on top of each other. A minimal number of gaps are introduced in the sequences so that the number

of matches is maximized according to a given optimization function.

In the past year, we ported a multiple sequence alignment program, developed by M.P. Berger and P.J. Munson, named MUSEQAL from the IBM PC to our Intel® iPSC/860 parallel computer and to the UNIX® workstation. In the coming year, in addition to evaluating this program, we will develop new multiple sequence analysis methods on the parallel computer.

Human Genetic Linkage Analysis

Recent advances in DNA sequencing technology made it possible for biological scientists to rapidly generate genetic markers for a particular set of families. This has resulted in a large amount of genetic data available for analysis. Human genetic linkage analysis uses statistical methods, based on inheritance patterns of these genetic markers in a chosen set of families, to draw genetic maps, estimate genetic risk, and identify genetic diseases. For example, genes for diseases such as Alzheimer's disease, manic-depressive illness, Huntington's disease, and cystic fibrosis have been identified by genetic linkage analysis. This ability to identify genetic diseases will eventually lead to therapeutic interventions.

In the last two years, CBEL, in collaboration with L.R. Goldin (NIMH), has implemented LINKMAP on the DCRT massively parallel computer. LINKMAP, one of the important programs in genetic linkage analysis, is the program generally used to draw the genetic map. LINKMAP infers the most likely position of a new genetic marker by iteratively calculating its likelihood at a series of points along an existing genetic map. Depending on the number of points, the number, size, and complexity of the chosen set of families, and the number of markers, the analysis performed by LINKMAP often takes 20 hours on a SUN® SPARCstation, and may take as long as a week. Exploiting the DCRT massively parallel computer, the analysis time can be reduced significantly. For example, an analysis that took four hours on the SUN SPARCstation was reduced to less than an hour.

In FY94, CBEL regularly participated in the Washington Theoretical Complex Disease Consortium (WTCDC) to provide its expertise in parallel computing techniques to participants who are interested in applying high performance computing to analyze complex genetic diseases. WTCDC was organized and led by S. Ghosh of National Center for Human Genome Research. A major goal of this consortium is to bring researchers from different fields such as genetics, biology, chemistry, computer science, mathematics, statistics, and medicine to share knowledge or collaborate on research projects to study complex genetic diseases. These researchers are not only from Institutes within NIH but also from universities in the Washington metropolitan area such as Georgetown University, American University, Catholic University, and Johns Hopkins University.

Complex genetic disease is generally defined as a genetic disease that is caused by more than one gene. To identify a complex genetic disease, a more sophisticated analysis is used than for a simple genetic disease. As a result, the computation is also more intensive. The complex genetic disease analysis can benefit greatly from high performance computing. In FY95, CBEL will continue to provide its expertise in high performance computing techniques to the WTCDC.

Protein Folding Prediction

The protein folding prediction problem is concerned with obtaining the 3-D structure of a protein molecule when only its amino acid sequence is known. Understanding the three-dimensional structure of proteins is important to studying their function in living systems and designing new ones for biological and medical purposes. The amino acid sequences of proteins are being discovered at an explosive rate. However, experimental procedures for determining their three-dimensional structure, such as x-ray crystallography and nuclear magnetic resonance spectroscopy, are slow, costly and complex. A need exists for theoretical and computational techniques that can be used to predict the structure from the sequence.

The protein folding problem remains unsolved because not all the biochemical rules that govern the folding and stability of proteins are known. If these rules were known, a computer program could be written to simulate the folding of a protein. In conjunction with the scientific work that is being done to understand the forces involved in protein folding, an alternative, computer approach is to write a program that searches through all the possible protein conformations in an attempt to find the ideal one. However, a search through the entire conformation space would require a prohibitive amount of computer time. The popular lattice-space Monte Carlo method has been used to significantly reduce the number of possible protein conformations by confining a protein onto a lattice. This method does not always yield a useful result because of the distortion introduced in the predicted structure by reducing the possible number of conformations.

At NIH, we are developing parallel computing methods for simulating protein folding so that more conformations can be considered and a more realistic energy function computed. This work involves strategies for searching through a large number of possible structures representing different energy states. The computationally intensive parts of a simulation are the long search through the great number of possible conformations and the computation of the free energy of the structures being considered during the simulation. We use a potential energy function that is the sum of three energy terms. The first term is an effective potential based on the phi, psi angle probabilities of the protein's main-chain. The second term is due to hydrogen bonding between the residues of the protein. The time to calculate these two terms is short. The third term is the hydrophobic potential, which is proportional to the solvent accessible surface area (ASA) of the protein molecule. The calculation of the ASA of the numerous structures considered in the simulation requires parallel computer performance.

In collaboration with B. Lee of the National Cancer Institute, CBEL is developing a computer program based on the dihedral angle space Monte Carlo procedure to simulate and capture protein folding events. The dihedral angle space Monte Carlo (DASMOC) procedure combines the use of knowledge-base rules and global minimization of the free energy. In essence, DASMOC samples and searches the conformational state space using the probability biased Monte Carlo techniques, with the hope that one or more good quality protein structures close to the global minimum of free energy will be found. DASMOC is a theoretical technique for simulating and capturing the protein folding events by performing global energy minimization. DASMOC represents a protein as a rigid chain that freely rotates about the axis of the peptide bonds connecting the central alpha carbon atoms with the amino nitrogen atoms and carboxyl carbon atoms. A single atom approximates the effect of each side-chain. Global energy minimization is implemented as a discrete optimization problem; it can be formulated as a search problem of finding an optimal protein structure with the lowest free energy. DASMOC searches the large and complex energy landscape to find a protein structure in the most stable energy state.

Because of the complexity of the task of searching through all the possible conformation structures and of the CPU intensive numeric computations, a computational solution to the protein folding problem requires an enormous amount of computer time to simulate and capture folding events. The goal of this project is to devise an efficient parallel search algorithm and other necessary parallel computing algorithms for an efficient implementation of DASMOC on massively parallel computers. In FY93, CBEL started work in developing a parallel search algorithm that can be used with the DASMOC procedure. It appears that a best first search augmented with heuristic pruning may satisfy the search requirement in DASMOC. CBEL also continued to develop parallel techniques for calculating the ASA of a protein. This effort will continue in FY95.

Nuclear Magnetic Resonance Spectroscopy and X-ray Crystallography

This activity involves the development of parallel software tools for NMR spectroscopy and x-ray crystallography. This includes tools for the three-dimensional structure determination and its refinement of biomolecules, using crystallographic data or NMR data.

In the past year, CBEL continued to support Frank Delaglio of the NIDDK Laboratory of Chemical Biology in his development of a parallel Genetic Algorithm (GA) approach to spectral assignment, determining which signals in the NMR spectrum belong to which atoms in the protein. CBEL also worked on the implementation of a parallel version of the X-PLOR program system, developed by Axel Brunger of Yale University, and widely used by x-ray crystallographers and NMR spectroscopists throughout the world. In the coming year, work will continue on this large software development effort.

Quantum Chemistry

The goal of this project is to develop *ab initio* quantum mechanical methods for use on massively parallel computer architectures. Unlike empirical force field (or molecular mechanics) or semi-empirical methods, *ab initio* methods are not parameterized; thus, they may be used to describe previously unknown chemical systems with a good degree of accuracy. Unfortunately, however, *ab initio* methods are quite computationally expensive; thus, to date, they have been applied only to small chemical systems (usually fewer than 30 atoms). In order to treat systems of biological interest (greater than 100 atoms), computers with speeds at least in the gigaflops (billions of floating point operations per second) will be needed. At present, massively parallel architectures provide the greatest hope of achieving the required speed economically.

Work to date has centered on implementing the Hartree-Fock Self-Consistent Field (SCF)

approximation to the time-independent Schroedinger equation on multiple instruction stream, multiple data stream (MIMD) distributed memory parallel computers. In the SCF method, the molecular wave function is described by a determinant of single-electron functions, known as orbitals, which are themselves expanded in terms of a set of known functions (basis functions). The potential energy term arising from electron-electron repulsion is treated by calculating an effective field due to the average positions of the electrons. This field is varied in an iterative fashion until self-consistency is reached. The computational bottle-neck in an SCF calculation is the construction of the Fock matrix, which depends on the calculation of O(n^4) (where "n" is the number of basis functions) electron repulsion integrals (ERI's). In the traditional SCF approach, these ERI's are calculated once, written to disk, and then read back in every SCF iteration. However, this requires a great deal of disk space. An alternative method, the direct SCF method, obviates the need for large amounts of disk space by recalculating the needed ERI's every SCF iteration. Although the direct SCF method seems to be much more expensive than the traditional approach, it is possible to greatly reduce the number of ERI's to be evaluated each iteration. Furthermore, since the information necessary to calculate all ERI's can be stored in memory on each node of a parallel computer, the direct SCF calculation may be easily parallelized.

The culmination of the first phase of a collaboration with Curtis Janssen and Michael Colvin of Sandia National Laboratories has been the development of a set of libraries for use in *ab initio* methods, as well as a prototype program, "mpqc", which makes use of these libraries. The programming language C was chosen for these libraries, since it allows for a great amount of flexibility and portability. Current capabilities of mpqc include closed-shell and high-spin open-shell SCF energies and analytic first derivatives, Mulliken and Lowdin population analyses, and electrostatic potential determination. Molecular symmetry is used to reduce the cost of both the energy and gradient calculations. Minimum and transition state searches may be performed in both Cartesian and internal coordinates. Most important, however, is the ability to use distributed matrices, greatly increasing the size of calculation that can be performed. If a complete copy of each matrix had to be held on each node, the size of problem that could be treated would be determined by the amount of memory on each node, regardless of how many nodes were used. By distributing the matrices, the size of problem that can be treated is determined by how many nodes one has. To our knowledge, mpqc is the only quantum chemistry package with this distributed matrix capability, and using this capability, we have performed SCF calculations on systems with as many as 2300 basis functions.

While single point SCF calculations are of some use, particularly in the determination of atomic point charges, what one most desires from *ab initio* methods are optimized structures. Only optimized molecular geometries are of use in the determination of most other molecular properties, as well as the energetics of chemical reactions. Given the great expense of each individual SCF calculation, it is imperative to optimize the geometry of a molecule in the fewest number of steps possible. To this end, very powerful optimization methods have been implemented in mpqc. Rather than being written in C, however, these methods have been developed in the object-oriented language C++. The usefulness of these methods can be demonstrated with one example from the scientific literature. The optimization of the molecule 7-thia-1,3-diazabicyclo (3.0.0) octa-2,4-dione ($C_5H_6N_2O_2S$, Cambridge Structural Database designation ACTHCP) is a good benchmark of the effectiveness of an optimization method. Using the simplest set of internal coordinates, Baker and Hehre were unable to optimize the structure of ACTHCP in under 100 geometry iterations. Using their new cartesian coordinate methods, Baker and Hehre were able to optimize ACTHCP in 90 iterations. In 1992 Fischer and Almlof, using a set of symmetry adapted internal coordinates as well as an empirically determined initial guess hessian, were able to determine the optimal geometry of ACTHCP in 25 iterations. In the same year, and Fogarasi, Zhou, Taylor, and Pulay were able to do it in only 13 iterations. Using an older incarnation of mpqc, which employed a conjugate gradient optimization method, we were unable to optimize the structure of ACTHCP in under 100 iterations. However, using the present mpqc code, we were able to determine the equilibrium geometry of ACTHCP in only 10 iterations. For a system of interest at the NIH, the disaccharide cellobiose, the conjugate gradient method previously available with mpqc could not converge on a geometry after more than 60 iterations. The internal coordinate optimization method recently added to mpqc, however, was able to optimize the structure of cellobiose in only 16 iterations.

The geometry optimization classes are the first product of phase two of the ongoing collaboration with Sandia National Laboratories. The primary thrust of this second phase of development will be the conversion of the existing mpqc libraries from C to C++. We feel

that only by using an object oriented language will we be able to provide a package with the requisite power, flexibility, and ease of use. Further, the classes we develop will be suitable for use in all areas of computational chemistry, not just *ab initio* quantum chemistry. In particular, we hope to explore density functional theory in the upcoming year.

Molecular Dynamics

CBEL continues to support the development of scalable parallel implementations of widely used molecular dynamics software packages, allowing for better levels of theory (e.g., electronic polarization), longer simulations giving better statistics, and larger molecular systems. This includes working with the Molecular Graphics and Simulation Laboratory of DCRT to support their parallel version of the CHARMM molecular dynamics program. This activity also includes collaborating with J. Saltz of the University of Maryland to exploit his data replicated version of CHARMM.

System Development

DCRT's Intel® iPSC/860 installation is somewhat unique, as this system is used for a wide variety of biomedical problems and by a sizable number of people with varying degrees of technical skill. In order to support the widespread use of the this parallel supercomputer in NIH's distributed computing environment, CBEL has implemented both hardware and software to aid in the use and administration of the system.

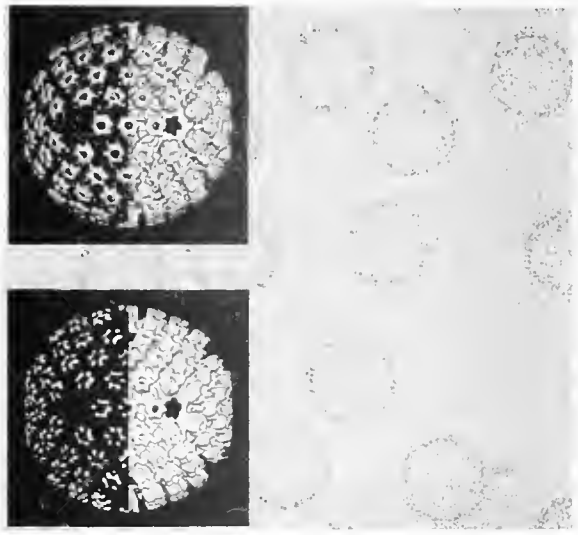
In FY94 CBEL acquired a number of SUN® SPARCstation 10 workstations along with a SPARCserver 1000 fileserver and over 30 GB of disk storage. These were immediately put to use, providing the auxiliary computing power and on-line storage necessary for researchers around the NIH campus, and elsewhere via the Internet, to access DCRT's parallel supercomputer. During FY94, CBEL also acquired and deployed an automated tape back-up system, to provide system administrators with a way to restore user data and software lost either due to user's own errors or in the event of a disk failure. The iPSC/860 was moved into the machine room of building 12 to provide a more stable environment.

CBEL also made software advances in FY94 to help researchers make use of the iPSC/860. A new module was added to CBEL's graphical front end, connecting to the PBatch batch scheduling system, as was a program that provides users with a convenient way to tell why certain jobs are running while others are still queued. As part of another project, work continued on the software needed to allow direct access to the parallel computer via a Macintosh personal computer.

Structural Biology: Image Processing of Electron Micrographs

B.L. Trus, Ph.D.
with A.C. Steven, Ph.D., E. Kocsis, Ph.D., F. Booy, Ph.D., J. Conway, Ph.D., A. Makhov, Ph.D., M. Cerritelli, Ph.D., J. Caston, Ph.D. (LSB/NIAMS); J.T. Shiller, Ph.D., R.Roden, Ph.D. (LCO/NCI); J. Brown, Ph.D., W. Newcomb (University of Virginia, Charlottesville, Virginia); T.S.Baker, Ph.D. (Purdue Univ., West Lafayette, Indiana); F. Homa (Upjohn, Kalamazoo, Michigan).

This project uses image processing techniques to analyze electron micrographs. To answer important questions in structural biology, it is necessary to obtain



Three-dimensional reconstruction of large icosahedral viruses. Shown are images of herpes simplex virus type 1 capsids, which illustrate the potential of this new technology. They show the location of a minor capsid protein called VP26 as mapped in experiments in which VP26 was first extracted from purified capsids by treatment with guanidine hydrochloride and then rebound to the capsids. The right-hand half of the top image shows the depleted capsid and the rebound VP26 capsid, and the left-hand half shows the 3D reconstruction, as it would be obtained with a conventional sequential computer. Parallel computing extended the analysis to obtain the bottom images, which improved the signal to noise ratio and the resolution from approximately 3.5 to under 3.0 nm. The clusters of six VP26 subunits, shown together in the top image, are clearly resolved in the bottom image (in collaboration with F.P. Booy, J.F. Conway, and A.C. Steven, NIH, and W.W. Newcomb and J.C. Brown University of Virginia).

relatively high resolution 2- and 3-D structural information about biological macromolecules. While atomic or near-atomic resolution information traditionally has been available by x-ray crystallography and NMR for some small molecules and proteins, the overwhelming majority of biological macromolecules are not crystalline, or are too large and therefore not amenable to x-ray crystallography or NMR.

Biological specimens can, on the other hand, be visualized in the electron microscope using a number of specimen preparation techniques. Negative staining and shadowing, which both use heavy metals, are two traditional approaches to increasing contrast to show the biological macromolecule's structure. Cryo-electron microscopy, a newer technique, attempts to preserve "native" structure by surrounding the specimen with a thin layer of ice. Collaborative studies with LSB, NIAMS are currently under way on a number of projects, whereby the electron micrograph images are computationally corrected, combined, averaged, reconstructed, or in some way computationally enhanced to improve the signal-to-noise ratio or to increase the interpretability of the structures being visualized. "Cryo" images are typically lower contrast and require greater computer processing to achieve satisfactory results.

Sometimes the image processing results can be combined with amino acid sequence analysis to yield additional information about the macromolecular structure. Sequence analysis uses the one-dimensional amino acid sequence of proteins together with both Fourier analysis and other predictive algorithms to attempt to identify parts of the sequence that may have a regular structure and to predict 3-D relationships.

Of particular interest to our research is the understanding of viral structures. At present we are continuing our efforts to investigate the structure of a large animal virus, human herpes simplex virus (type 1). We are completing the localization of the major capsid proteins and attempting to obtain higher resolution structures. Using the 3-D icosahedral reconstruction technique, we apply the symmetry of these virus particles both to finding the orientation of randomly oriented capsid particles (in ice) and to combining many particles into a 3-D reconstruction. Biological material for these herpes virus reconstructions is provided through a collaboration with researchers at the University of Virginia, Charlottesville and from the Upjohn Co. (Kalamazoo). The electron microscopy is performed in LSB, NIAMS. Interpretation of our 3-D reconstructions is performed jointly by all collaborators.

Starting with the precursor herpes capsid (B-capsids), we have studied degradation products (e.g., after guanidine HCl or urea treatment) with the goal of determining the 3-D location of the seven major capsid proteins. Higher resolution difference 3-D reconstructions, for example, clearly show that one protein, VP26, is bound on the outer tips of the hexons. In a related project, 3-D reconstructions of *baculovirus* produced herpes proteins (from Upjohn) have been used to produce reconstituted capsids (at the University of Virginia), which we have shown are identical to herpesvirus produced *in vivo*. By systematically combining different proteins, we hope to obtain additional information about the high resolution details of herpes virus capsid structure. In another, related research project, we have analyzed the structure of channel catfish virus, which has been found to be amazingly similar to its HSV-I cousin. We may be able to learn more about the evolution of herpes virus capsids from such studies.

Future work on this project involves the use of additional antibodies to confirm our localization experiments of other major proteins, and a continuing attempt to increase resolution. The computational demands of the 3-D reconstructions have prompted the use of DCRT's iPSC/860. This year, progress has been made in combining many more images in 3-D reconstructions, in part due to the availability of the Intel® supercomputer (see the section on High Performance Biomedical Computing). A new software program (EMPFTREF), provided by T.S. Baker (Purdue University), promises of more accurate orientation refinement, and therefore high resolution 3-D reconstructions.

A number of other collaborative projects in structural biology are currently in progress. We are using similar 3-D reconstruction techniques to study the structure of icosahedral L-A virus (from yeast), papillomavirus, and poliovirus. We have compared the structures of full (RNA containing) L-A virus with empty L-A virus. In a new study of papillomavirus (in collaboration with NIAMS and NCI), we have verified the known structure of bovine papillomavirus (bpv), and have recently obtained a new 3-D reconstruction of antibodies to the L1 protein of bpv. We hope to be able to localize the two major proteins of bpv, as well as to understand more fully the function and activity of a number of papilloma antibodies. In a related methodological study, we have examined the effects of radiation dose on cryo-electron microscopy, and the resolution obtained from 3-D reconstructions as a function of dose. These results have recently been published. A 2-D project involves a study of the

connector proteins FHA from *Bordetella pertussis*. In this study, the amino acid sequence data has been combined with image processing results to yield additional useful information about the macromolecular structure. A report of this work has been accepted for publication.

Biomedical Image Processing

B.L. Trus, Ph.D. and M. Vivino
with D. Chow, K. Kempner, W. Gandler, (DCRT/DSB); W. Rasband (NIMH/MHIRP); M. Datiles, MD., A. Mahurkar, L. Lopez, MD, B. Magno, MD, (NEI/OGCS); G. Makara (NICHD/DEB); S. Yoshikami, W. Hagins (NICHD/LCP); R. Hewitt (NCI/DCBDC); D. Adams, Ph.D. (IIT/Technion)

This project uses sophisticated image processing techniques to analyze biomedical images. The goal is to establish collaborations with biomedical experts who require new algorithms and possibly new hardware capability to solve difficult imaging problems. Typically, complex new mathematical algorithms, as well as new combinations of existing algorithms are utilized. We attempt to apply the best computer platform for each problem, using such diverse computers as an Apple Macintosh®, a DEC VAX® or Alpha®, a SUN® workstation, or an Intel® iPSC/860 supercomputer.

Two current projects include ophthalmic image analysis and general consulting to the NIH scientific community in biomedical image processing. In collaboration with the National Eye Institute, we continue to develop systems to quantitate lens opacities (cataracts) and assist in diagnosis of ocular diseases. For cataract studies, the computer assisted instrumentation allows observation of the effects of anti-cataract drugs or routine pathological grading. During the last year we completed a system that analyses retro-illumination images. This device projects light onto the retina and then captures an image of the lens with reflected light. The technique of reflecting light from the retina does not always produce a perfect image, and sometimes leaves a distortion pattern in the image of the retina. While this distortion limits the device's effectiveness, it is the best system available to evaluate the anterior and posterior subcapsular cataracts. Before making quantitative morphological and densitometric measurements on these images, our software removes the distortion pattern. Data on the software was presented at the Advanced Research in Vision and Ophthalmology 1994 conference and has been submitted for publication.

An integral part of our image processing consulting is ongoing support for the NIH Image program (developed by Wayne Rasband/NIMH/MHIRP). Our support includes continuing development of new algorithms and four supporting documents, which are now distributed with the package. These documents are widely used and referenced both in the intramural program and by extramural biomedical scientists. These documents include: a guide on how to modify source code, which is intended to help scientists develop new user applications or macros; a technical guide describing scientific application usage with the package; a list of frequently asked questions and answers; and support for a guide authored by David Chow of DSB concerning analysis of gels.

We are also offering support in the development of specialized image processing applications for campus users. These applications involve methods to provide repetitive analysis automation, video acquisition, quantitative analysis, counting, particle location, identification, etc. Our support is tailored to the individual needs of a researcher and primarily involves developing an application not generally available in any of the commercially available image processing packages. For example, this year a set of routines were developed for specialized video acquisition requirements. The routines automatically integrate sequential frames of video, measure a user-defined, irregularly shaped region, then repeat the process automatically and continually for hours. In another instance, software was developed for a user who wished to determine whether granular particles in her media occurred periodically. For this purpose an algorithm was used to identify the individual particles, find the particles' location in a Cartesian coordinate system, and produce a distance matrix for statistical analysis. We also examined image processing routines for applicability in fluorescent analyses in microscopy, routines to translationally and rotationally align sections for the 3-D reconstruction of viscous pathways in a macrophage, routines to automatically copy portions of images into a larger montage for publication and reference, and routines to copy these same portions into a stack for automated analysis. Hardware and software support was given to scientists wishing to control a motorized stage during video acquisition, import the image files into a different standard formats, and acquire video with the new AV Macintosh.

Future projects include collaborating in the development of computer systems to analyze light microscopy images (including performing real time 3-D reconstructions), as well as analysis of images from PET, SPECT, and MRI, for example. We anticipate requests for collaboration in other high technology biomedical imaging projects.

Publications

Adam DR, Kempner KM, Vivino MA, Tucker EE, Jones M. Correction of arterial flow velocity Doppler color mapping in transverse cross-sectional imaging. In: Computers in Cardiology. Los Alamitos, CA: IEEE Computer Soc Press, 1993; 281-284.

Adam DR, Kempner KM, Vivino MA, Tucker EE, Jones M. Measurement of flow velocity in a phantom by corrected Doppler color mapping technique, using transverse cross-sectional imaging. In: Computers in Cardiology. Los Alamitos, CA: IEEE Computer Soc Press, 1993; 277-280.

Booy FP, Trus BL, Newcomb WW, Brown JC, Conway JF, Steven AC. Finding a needle in a haystack: detection of a small protein (the 12 kDa VP26) in a large complex (the 200-MDa capsid of herpes simplex virus). Proc Natl Acad Sci USA 1994; 91:5652-5656.

Booy FP, Trus BL, Newcomb WW, Brown JC, Conway JF, Steven AC. Localization of the 12 kDa Protein, VP26, in the Capsid of Herpes Simplex Virus Type 1. In: Proceedings of the 13th International Congress on Electron Microscopy. 1994;1:27-28.

Carson RE, Yan Y, Chodkowski B, Yap TK, Daube-Witherspoon ME. Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm. IEEE Trans. on Medical Imaging 1994;13:526-537.

Conway JF, Trus BL, Booy FP, Newcomb WW, Brown JC, Steven AC. The effects of radiation damage on the structure of frozen hydrated HSV-1 capsids. J Struct Biol 1993;111:222-233.

DasSarma S, Lanczycki CJ, Ghiasas SV, Kim JM. Defect formation and crossover behavior in the dynamic scaling properties of molecular beam epitaxy. Phys Rev B 1994; 49:10693.

DasSarma S, Ghaisas SV, Lanczycki CJ. Dynamical correlations in activated growth models of molecular beam epitaxy. Phys Rev E (in press).

Johnson CA, Weisenfeld NI, Trus BL, Conway JF, Martino RL, Steven AC. Orientation determination in the 3-D reconstruction of icosahedral viruses using a parallel computer. In: Proceedings of Supercomputer '94, 1994 (in press).

Johnson CA, Yan Y, Carson RE, Martino RL, Daube-Witherspoon ME. A system for the 3-D reconstruction of retracted-septa PET data using the EM algorithm. In: Proceedings of the 1994 IEEE Medical Imaging Conference, 1994 (in press).

Kocsis E, Trus BL, Steven AC, Smith PR, Hannah JH, Brennan MJ, Kessel M. Orientation of porin channels in the outer membrane of Bordetella pertussis. Mol Microbiol 1993;9:469-476.

Lanczycki CJ, DasSarma S. Noncquilibrium influence of upward atomic mobility in one-dimensional molecular beam epitaxy. Phys Rev E (in press).

Lanczycki CJ, DasSarma S. Dynamics of step roughening in vicinal surfaces. Phys Rev Lett (in press).

Makhov AM, Hannah JH, Brennan MJ, Trus BL, Kocsis E, Conway JF, Wingfield PT, Simon MN, Steven AC. Filamentous hemagglutinin of *Bordetella pertussis*: Monomeric hairpin model of a bacterial adhesin. Biophys J 1994; 66:A341.

Makhov AM, Hannah JH, Brennan MJ, Trus BL, Kocsis E, Conway JF, Wingfield PT, Simon MN, Steven AC. Filamentous hemagglutinin of *Bordetella pertussis*: A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. J Mol Biol 1994;241:110-124.

Martino RL, Johnson CA, Suh EB, Trus BL, Yap TK. Parallel computing in biomedical research. Science 1994;265:902-908.

Martino RL, Johnson CA, Suh EB, Trus BL, Yap TK. The role of high performance parallel computing in biological research. In: Proceedings of the IEEE Engineering in Medicine and Biology Society Sixteenth Annual International Conference, 1994 (in press).

Trus BL, Davison AJ, Booy FP, Steven AC. Capsid structure of channel catfish virus, and evolutionarily distant herpesvirus, by cryo-electron microscopy. In: Bailey, G.W. ed. Proceedings of the 52th Microscopy Society of America, 1994;114-115.

Trus BL, Homa FL, Thomsen DR, Newcomb WW, Booy FP, Brown JC, Steven AC. Visualization of herpesvirus capsids assembled in a recombinant baculovirus system by cryo-electron microscopy. In: Proceedings of the 13th International Congress on Electron Microscopy, 1994;3A:527-528.

Vivino MA, Chintalagiri S, Trus B, Datiles M. Development of a Scheimpflug slit lamp camera system for quantitative densitometric analysis. Eye 1993; 7:791-798.

Vivino MA, Maharkar A, Trus B, Lopez L, Datiles M. Quantitative analysis of retroillumination images. Eye (in press).

Yap TK, Frieder O, Martino RL. Parallel computing in biomedicine: genetic and protein sequence analysis. In: Zomaya AY, ed. Handbook of parallel and distributed computing. New York: McGraw-Hill Book Company, 1994 (in press).

Presentations

Johnson CA. Orientation determination in the 3-D reconstruction of icosahedral viruses. Presented at: NIH Research Festival. Bethesda, September 1994.

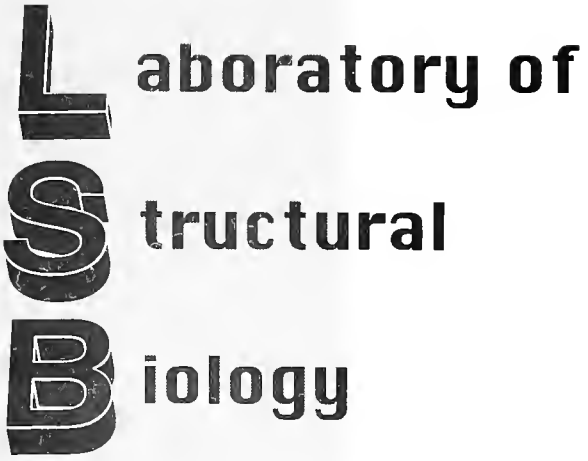
Martino RL. Parallel computing in structural biology and medical imaging. Presented at: The George Washington University Department of Electrical Engineering and Computer Science Spring 1994 Colloquium Series. Washington, D.C., March 1994.

Martino RL. High performance parallel computing in a biomedical research environment. Presented at: The ARPA High Performance Computing and Communications Symposium. Alexandria, VA, March 1994.

Martino RL. High performance computing in biomedical research. Presented at: The FGIPC Eighth Annual High Performance Computing and Communications Conference. Research Triangle Park, NC, June 1994.

Martino RL. Making high performance parallel computing part of the biomedical research process. Presented at: The 1994 Intel Supercomputer Users Group Conference. San Diego, CA, June 1994.

Trus, BL. Image processing in electron microscopy. Presented at: NIH Research Festival. Bethesda, September 1994.


Laboratory of
Structural
Biology

Laboratory of Structural Biology

V. Adrian Parsegian, Acting Chief

After a merger of four different research groups concerned with questions of molecular structure and organization, this Laboratory has now enjoyed its first full year as a distinct entity.

The *Section on Molecular Forces* headed by V. Adrian Parsegian, focuses on the measurement and explanation of forces between and within macromolecules. Its strategy is to use intermolecular forces to understand the molecular recognition and specificity. The same "osmotic stress" method used to measure forces between molecules is also being applied to see the forces within molecules undergoing changes in structure. Inter-institute collaboration with NIDDK, and a sharing of laboratory space and research partners, provides an essential connection between computer research and "wet-laboratory" science.

Protein folding is the central theme of computations by Richard Feldmann building on his earlier success in molecular graphics with his ability to link distant systems into cooperative computation. Rather than the frontal assault wherein one tries to fold a protein all at once, the approach here is to compute stable structures on pieces of the protein as it emerges during synthesis.

The *Analytical Biostatistics Section,* under Peter Munson, tests and devises statistical methods for a large range of biological systems—as diverse as molecular structure prediction and bone growth. The group has developed a new protein secondary structure prediction method, which makes use of multiply aligned sequences. The LIGAND and ALLFIT programs, developed and supported by the Section, continue to be used widely.

The *Molecular Graphics and Simulation Section* is widely known for its work on the popular CHARMM molecular dynamics program. This group develops new methods for macromolecular simulation and for modeling complex systems. Its activities this year have also included establishment of a coordinated cluster of high-performance workstations. The group has also enjoyed success with computations on HIV protease and myoglobin.

The joint development with CBEL of a high performance parallel supercomputing system "to solve important computationally intensive problems in basic and clinical biomedical research" was recognized in a NIH Director's Group Award in June.

The force measurement strategy has experienced rapidly spreading use—from the identification of osmotic stress as causing loss of specificity in DNA restriction enzymes (S. Sligar, Dept Chemistry, University of Illinois), to measuring hydration changes during substrate/enzyme binding (P. Rand and P. Nicholls, Brock University, and J. Kornblatt, Concordia University, Canada), to the study of industrial polysaccharides (B. Cabane, CNRS, France), to a method for making ceramics (C. Zukoski, Dept Chemical Engineering, University of Illinois), as well as in many basic research applications. Parsegian's popular course, "Physical Forces Organizing Biomolecules", was presented under the aegis of the Inter-institute Structural Biology Interest Group, of which several LSB members are active participants.

Section on Molecular Forces

V. Adrian Parsegian, Head

Direct Measurement of Forces between Membranes or Macromolecules

The theme of this work is to develop a useful, accurate science of the forces that organize biomolecules. To this end, we have accelerated our efforts to measure forces between proteins, DNA double helices, and polysaccharides. We have also concluded a set of studies on the release of water upon DNA/ protein and DNA/drug binding. The Osmotic Stress method, developed by members of this group for direct force measurements, has enjoyed widespread use in dozens of other laboratories around the world, a remarkable "technology transfer." The growing catalog of information about interactions continues to create a new logic for thinking about molecular recognition, and folding.

Forces Between Proteins

Forces between collagen triple helices have been measured under different solution conditions, and separate repulsive and attractive contributions to the interaction have been determined. Within experimental error, the same, purely exponential repulsive force curve is observed with:

- pH lowered from 7.5 to 6
- the presence of glycerol or glucose
- reduction of collagen by sodium borohydride
- slow aging of fibers stored over a year in PEG solution.

This exponential force is the repulsive part of the total interaction. We find the attractive forces

responsible for self-assembly of collagen fibers by subtracting this repulsion from the net force measured under normal conditions at pH 7.5. The resulting attractive force-distance curves are also exponential, but they decay twice as slowly as the repulsion. The attraction shows unusually strong temperature dependence. From 5°C to 35°C it increases by a factor of two. It also shows unusual sensitivity to small hydroxyl compounds. The attraction can be disrupted by glycerol and glucose, but not by methanol, ethylene glycol and sorbitol.

We have found that, while the attractive forces between the triple helices are strong enough to induce self-assembly of collagen fibers, an extra component is required for more tight packing of collagen in tendons. The exact nature of this component has not yet been established.

We have started preliminary measurements of forces between tropomyosin molecules, the first step in applying the osmotic stress force measurements to muscle proteins. We plan to expand these measurements to learn more about protein-protein interactions.

Building on previous observations, we have now dissected different contributions to forces measured between collagen molecules. This study is already providing important practical information on the assembly and function of collagen in living tissues, under normal conditions and in disease. It has driven us to reevaluate the existing theories of protein assembly and folding and to suggest new theoretical models, which are now undergoing extensive testing.

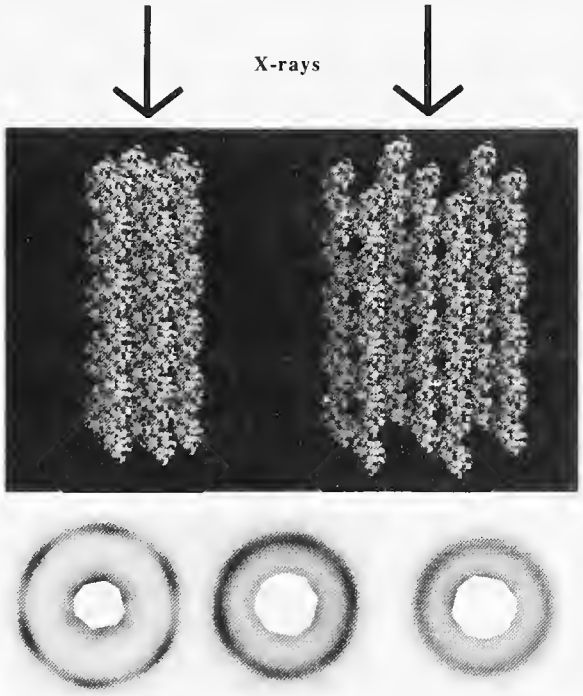
It has been almost axiomatic that temperature-induced protein folding and assembly reactions are dominated by the hydrophobic effect, the entropically favorable removal of water structured around exposed non-polar residues. Now, actually measuring intermolecular forces, we can better discriminate between the contributions of hydrophobic and hydrophilic interactions. From the sensitivity of the attractive forces between collagen molecules to pH, carbohydrates, and alcohols and from the magnitude of the temperature dependence we conclude that the attraction is caused by interaction between hydrophilic residues. Surprisingly, the contribution from the hydrophobic effect appears to be negligible, at least in the case of collagen. The titration experiments and measured specificity to small hydroxylic compounds suggest that the interhelical attraction is mediated by a hydrogen bonding network in interstitial water that originates from histidine residues.

This work is intended to provide a vocabulary of forces to be incorporated into computer algorithms for protein folding, assembly, and ligand or drug binding.

Forces Between DNA Double Helices

This year, we published the first of our intended "toolbox" papers, which codify DNA-DNA measured forces in a form that can be used in computation and analysis of molecular assembly. These forces are themselves the center of our own investigation into the packing and packaging of DNA into ordered assemblies, such as in viruses.

The osmotic stress method used to measure electrostatic, hydration and configurational forces is now being used to measure the work of changing the over-all packing of DNA double helices from the liquid-crystalline arrays seen in nature to various, looser arrangements. As before, we see the concerted action of



X-ray diffraction of condensed hexagonal DNA arrays. The left diffraction ring is at an interhelical spacing of about 25 Å, and the right one at about 40 Å. As the density of DNA is increased (smaller DNA-DNA interhelical spacing), the diffraction ring assumes hexagonal symmetry, indicating the onset of long-range bond order, while the positional order remains short-range. This means that, although the positional correlation in the plane normal to the average orientation of the helices are almost liquid-like, the correlations in the bonding structure, i.e., the number of nearest neighbors and their relative orientation, are long-range. This is the first instance of measurement of bond order in condensed DNA arrays (or polymers arrays in general).

electrostatic double-layer repulsion working against molecular disorder. Different salts and different neutral solutes influence intermolecular forces to effect changes in DNA packing. For the same reason, different arrangements of DNA show a preference for different ions or neutral solutes.

The importance of DNA interactions, and the delineation of the relatively simple rules by which they pack in viruses, and of the more specific ways they are packaged in cells, is beyond argument. DNA in solution and in molecular assemblies has taken on another life, as an ideal object for polymer theorists to learn the rules that decide why a molecule stays in solution or condenses into different forms (nucleosomes, chromosomes). In either case, the measured forces and energies incurred in packing provide critically important information.

DNA/Protein and DNA/Drug Interactions

This year has seen the first quantitative measurement of the amount of water released upon specific vs. non-specific binding of DNA to protein (lac repressor) or upon the binding of DNA to various drugs. There is an immediate energetic connection between these changes in molecular hydration and the powerful "hydration forces" measured between large molecules brought into contact.

Specifically, direct measurement of forces between molecules in condensed arrays show a dominating contribution from water structuring. Now measurements of the osmotic sensitivity of DNA-protein and DNA-drug complex formation in dilute solution are also showing a link between binding strength and structured water release. Binding constants for the specific association of galactose repressor with operator differ from those for nonspecific binding by a factor of 10^4. This difference is reflected in the release of some 150 more water molecules for operator binding than for nonspecific association. Much smaller binding energy differences of repressor to different operator sequences are even seen as differences in water release. The binding constant of the gal repressor-O_E operator association is twice that of the repressor-O_I complex and is accompanied by the release of 6 more water molecules. A connection between energy and water release is also seen for drug-DNA interactions. The binding constant of netropsin to an Eco R1 site is 10 times stronger than to an Nde 1 site and binding also releases some 30 more water molecules.

An important role for a change in hydration accompanying a nucleic acid conformational transition has been uncovered. The sensitivity to water activity of the B-Z transition of DNA indicates that a difference in the number of solute excluding water molecules between the two conformations is central to the transition. Conditions in the cell may be sufficiently crowded to allow the Z-form to exist without other stresses.

We will proceed systematically to collect primary data correlating changes in hydration with the energetics and thermodynamics of the interaction of sequence specific DNA binding proteins and drugs with DNA. Of particular interest is the osmotic sensitivity of Cro repressor binding to different sequences. The binding of Cro repressor protein to DNA has been particularly well characterized thermodynamically. Free energies, enthalpies, entropies, and heat capacities have been determined for a set of sequences that span a wide range of binding constants. The linear dependence of enthalpy and entropy is consistent with a hydration compensation and with our previous observations of Mn^{2+}-DNA assembly.

The relative osmotic sensitivities of binding to the different sequences will provide a key link between the thermodynamics and the underlying physics of recognition. The demonstration of hydration forces dominating the interaction of molecules in solution will fundamentally change the way one thinks about binding and energetics. An effort will be initiated to extend the osmotic stress technique to eukaryotic transcription complexes.

Eukaryotic gene activation depends more on protein-protein interactions than for prokaryotes. As yet, these interactions can only be indirectly inferred. The osmotic stress technique offers the opportunity not only to measure the extent of protein-protein contact, but also to stabilize these relatively weak complexes for further physical characterization.

Hexagonal-Lamellar-Hexagonal Re-entrant Phase Transition in Phospholipid Membranes

Our theoretical work has included a quantitative theory of transitions between lamellar and non-lamellar forms of phospholipids. The strains in molecular packing that are seen in these transitions are known in at least one case to be related to the strains in peptide-lipid interactions that change the probability of forming ionic channels.

In collaboration with M. M. Kozlov from Free University Berlin, Germany, and R. P. Rand from Brock University, Ontario, Canada, we have accounted for the unusual structural change when dioleoyl-phosphatidyl-ethanolamine (DOPE) undergoes a hexagonal-lamellar-hexagonal transition sequence as the

water content is reduced. We describe the role played by the energies of bending, hydration, voids in hexagonal interstices, and van der Waals interaction in this transition sequence. We have used x-ray diffraction and osmotic stress experiments on the two phases to derive the structural parameters and all of the force constants defining the energetics of the hexagonal and lamellar phases. We have calculated the chemical potentials of lipid and water in both phases and derived the phase diagram of the lipid with no free, adjustable parameters. The calculated temperature/osmotic stress and temperature/composition diagrams quantitatively agree with experiment. The re-entrant transition appears to be driven by a delicate balance between the hydration energy in the lamellar phase and bending energy in the hexagonal phase, while the energy of voids in hexagonal interstices determines the energy scale and temperature range. Van der Waals attraction between the bilayers in the lamellar phase does not appear to be important in this transition.

Lamellar to hexagonal phase transition is a central step in rearrangements of membrane lipids in cells. It is thought, for example, that nucleation of the hexagonal phase in cell lipid matrix is a necessary intermediate step in membrane fusion. This is the first quantitative description of this complex process of the hexagonal phase formation to account for individual contributions of different interactions between lipid molecules. It is also a new step toward understanding how intermolecular forces induce conformational/phase transitions in biological systems.

The "Long-Range" Hydrophobic Force

A second line of theoretical work concerns the origin of an enigmatic "long-range hydrophobic force" that has been reported between non-polar surfaces. This puzzling force seems to us explicable in terms of conventional electrostatic and van der Waals forces, allowing no justification for its *ad hoc* invocation in recent theories of polymer organization.

Even as they are invoked in practically every theory of lipid and protein organization, "hydrophobic forces" often provoke considerable confusion among theorists. The reported existence of strange forces of this type, but acting at distances of thousands of Å, has been an especially vexing phenomenon. That the evidence now can be interpreted in terms of easily understood electrostatics will clarify at least the most mystifying of these forces.

Future Plans

In addition to continued measurements on proteins and DNA, we will investigate forces between polysaccharides, particularly model systems such as guar, chitosan, and hydroxypropylcellulose. We have already begun a series of measurements on forces among these stiff polysaccharides, a most neglected category of bio-materials. The object here is to learn what governs the organization of large sugars that can aggregate into strands or sheets and then to interact in those forms. There is a strong technological as well as biological motivation for understanding these processes, because of their widespread occurrence and industrial use. There should be important biomedical applications in the understanding of extracellular matrix abnormalities of cartilage, bone, vascular basement membrane and connective tissue disorders.

We are also undertaking new theoretical studies on hydration forces. For example, we are investigating how these forces, due to perturbation of water structure by macromolecular surfaces, will change in concentrated salt solutions, where the charged ions also re-organize water. Concomitant with that work is a re-examination of the idea of a "dielectric constant" in water that is disturbed by a protein or a DNA surface.

Since we are now able to measure forces between proteins, and since urea is known to be such a powerful agent to change protein structure, we have begun to measure forces in the presence of urea, as well as of several other protein-active solutes. We have also made the first, promising measurements of the change in the vibrational spectrum of water near collagen; these measurements will be made on the same samples used for force measurement, so that the ideas of hydration, perturbation of water, and intermolecular force can be more instructively united.

Physics of Ionic Channels and Other Proteins with Aqueous Cavities

Our objectives here are:

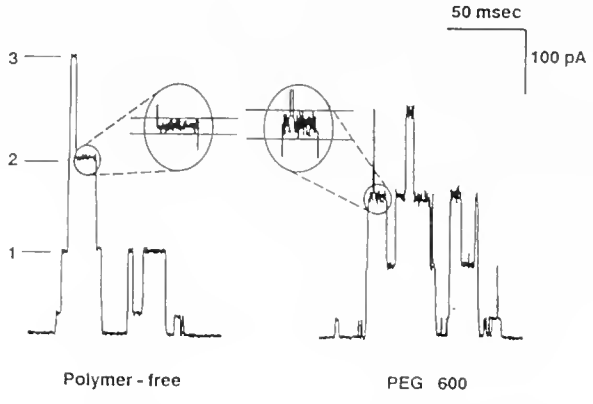
- to "interrogate" the structural and functional features of ionic channels by their reaction to polymers of varied size
- to use physical "noise" measurements to study kinetics of ionic channels; in particular, to resolve rapid events such as passage of small nonelectrolyte molecules and binding of ions at charged sites
- to develop practical strategies for improving

signal/noise ratios by signal mixing and computer analysis

- to test for channel-creating capabilities of antibiotics and to develop assays of antibiotic action by observations on model membranes
- to relate the forces measured between macromolecules to the energetics of protein conformation seen in the opening and closing of ionic channels.

Studies with Alamethicin

Channels made from the peptide antibiotic alamethicin have been observed while they are subjected to the osmotic action of differently sized neutral polymers. It is possible not only to see the degree of penetration of the polymers into the channel from their osmotic action but also to follow the kinetics of motion of small polymers through the ionic channel. When polymers enter the channel, there is a flickering of electrical current, whose frequency indicates the time of passage of polymers. It is thereby possible to measure not only the likelihood of polymers entering small spaces but also their diffusion constants.



Ion current of a single ionic channel as changed by polymer addition (Nature 1994;370:279-281). Polymer decreases the amplitude of different conductance levels (numbers 1, 2, and 3 at left) due to partitioning of polymers inside the channels and blocking current path. It also induces additional current noise because of random polymer exchange between channel and bulk solution.

While the original intention of our osmotic stress study was to determine channel structure and the energetics of channel formation from the effects of polymers, it has led us unexpectedly to be able to study one of the most fundamental problems in statistical physics, the "ergodic hypothesis". By learning how to watch and measure the passage of one molecule traversing one open channel, we learn that a protein is able to take an average of its surroundings over the time that it is in a particular state. Individual proteins are seen to act as individual statistical mechanical systems.

The ability to watch single molecules teaches us how to gauge the dimensions of protein cavities by the action of differently sized molecular probes. It also provides otherwise-unavailable information on the properties of the polymer-probes themselves, information of value to polymer scientists.

These channels are sensitive to the phospholipid identity of the bilayer into which they are incorporated; in particular, there is a strong correlation between the probability of high-conductance states and the tendency of the phospholipid to form non-lamellar structures.

The ability to see continuously modulated changes in membrane function with systematic variation in membrane lipid is a potent tool for addressing the question of the importance of lipid composition in cell development.

Studies with Roflamycoin

The Hofmeister effect is shown to apply to transport properties of ionic channels. Binding times, conductance effects and cationic selectivity of chaotropic anions on roflamycoin channels depend on their position in Hofmeister series.

Determining a connection between ion transit specificity and their binding to large molecules provides an opportunity to connect poorly understood transport properties with solution physical chemical properties. Here in particular, one can work on a connection between ion binding to channels and the effect of ion binding to proteins that is seen in our measurements of forces between collagen triple helices.

Two-Sided vs. One-Sided Formation of Channels by Amphotericin B

Many antibiotics work by perforating the infectious agent by ionic channels. At the same time, the host-cell membranes are also at risk of perforation by the same drug. The action of the anti-fungal agent Amphotericin B (AmB) is thought to be that the drug acts more efficiently on the in ergosterol-containing fungal cell than on the cholesterol-containing mammalian host-cell membrane. *In vitro* tests of comparative action must recognize that drugs are usually more concentrated outside a cell, and that AmB forms different kinds of channels when added only on one side of a membrane.

We have found that there is a strong sterol-based discrimination in one-sided AmB action on artificial phospholipid bilayers. This discrimination is different, however, from that seen when the drug is presented on both sides.

Artificial bilayer membranes can provide a useful test-bed for evaluating the action of membrane active drugs. Questions of asymmetric administration/ formation and sensitivity to membrane composition can be systematically addressed. Drugs can be screened by monitoring their potential deleterious effects first in such *in vitro* systems.

Future Plans

We will concentrate on the dynamics of ions and neutral solutes inside the ion channel pore. Specifically, we will study in more detail reversible ion binding to the residues of the channel-forming proteins and basic physical principles governing partitioning and kinetics of soft polymer molecules in the channel pore. We will explore the possibility of observing "stochastic resonance" in a system of voltage-dependent ion channels, studying information transduction in the presence of external calibrated noise sources.

Modeling the Mechanism of Protein Folding

Richard J. Feldmann

This year, computational folding of proteins moved from an abstract geometry-free model developed during the previous year to a 3-dimensionally embedded constraint evaluation model. The geometry-free model was implemented as a series of topological connections between charged atoms, representing the hydrophilic aspects of peptide biochemistry, and another series of connections between groups of carbon atoms, representing the hydrophobic aspects. The rule-based manipulation of topological connections is, computationally, relatively inexpensive. Simulation of protein folding using this rule-based model produces local secondary structure in the form of helices but can produce neither global beta sheet secondary structure nor correct global packing of the protein.

The topological connections between hydrophilic and hydrophobic aspects of protein structure can be represented as distance constraints between atom pairs in a constraint satisfaction program such as DGEOM© (developed by Jeff Blaney, now at Chiron Corp). Both x-ray crystallographic and NMR structure determination routinely use programs of this type. DGEOM© is relatively expensive to use because the size of the program expands as the square of the number of atoms in the protein, and the time to determine a trial structure also increases as the square of the number of atoms. In addition, we have found through the year that it is necessary to compute 16 to 20 trial structures in order to be sure that the pattern of constraints is optimally satisfied. CHARMM has also been added to the structure simulation process to bring each best trial structure to an energy minimum. This combined use of DGEOM©and CHARMM has increased the computation burden for simulation of protein folding by a factor of 40, using our methods.

In order to maintain a model that is physically reasonable, the sequential synthesis of the protein from N to C terminus is modeled. In ribosomal synthesis, it is only when the peptide emerges from the ribosomal that it begins to adopt a folded conformation. The linearly increasing peptide length keeps the DGEOM© computation time to a minimum. We have found that a pattern of hydrophilic and hydrophobic constraints develops with the lengthening of the peptide in such a way that the peptide adopts conformations that are very close to the crystal or NMR structure soon after simulated ribosomal emission. As the sequence for a helical portion of a peptide is emitted, it folds into a helix. As soon as the sequence for an anti-parallel beta sheet is emitted, it too folds into the correct secondary structure. The most striking result from this year's simulations is the discovery that when the strand-helix-strand peptide sequence emerges, it forms a tertiary structure with the correct macroscopic handedness.

Ten classes of protein architecture are being studied in parallel in an attempt to make the computational simulation of protein folding as general as possible. We have observed that, in all structural classes, the correct local secondary structure is formed, and often the correct macroscopic handedness is also formed. So far, the simulation program has not had success in reproducing the tightly packed atomic structure characteristic of crystal and NMR structures. Rules and parameters are continually added and modified in attempts to produce better packing.

Computational resources for our experiments in protein folding come from a variety of sites (the DCRT, NASA Lewis, Argonne National Laboratory, the Maui High Performance Computer Center and the Pittsburgh Supercomputer Center). By distributing the computational load over a number of sites and architectures, it has been possible to look for windows of opportunity on an hourly basis and to investigate the interaction between program structure and machine system architecture. We have found that the NQS batch queue at NASA Lewis provides a 5-fold increase in throughput when all ten classes of protein architecture are being simulated at once.

During this year, there has been a strong synergy between survey of the literature, biochemical reasoning, computer programming and computational resource

availability. Keeping folding experimentation to a day or two in length makes it possible to make several sets of experiments per week. In the next year we would hope to pass from the molten globule state that is characteristic of the type of atomic packing achieved today to the more compact structure state characteristic of crystal and NMR structures.

Analytical Biostatistics Section

Peter J. Munson, Ph.D., Head

Statistical and Computational Methods for Physiology, Pharmacology, Endocrinology and Molecular Biology

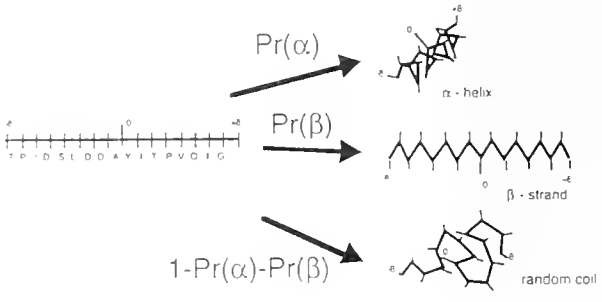
Newly developed computational, mathematical-statistical methods are frequently underutilized in many areas of biomedical research. Yet such methods may concisely address questions posed by researchers. Newly developed statistical techniques such as "cross-validation" and "bootstrap" can tax or even exceed the available computer resources, yet they promise nearly optimal solutions to questions of significance or interpretation of unique datasets. The purpose of this project is to investigate the applicability of such modern statistical methods to problems in a wide area of biomedical interest, and to implement and disseminate these applications. Further, where necessary, we seek to develop new methodologies specific to particular biomedical problems or datasets. In particular, we have applied these techniques to protein structure/function prediction from sequence, characterization of the database of protein structures, analysis of periodicity in long DNA sequences, analysis of characteristics of growth curves, analysis of pulsatile hormone secretion, and modeling of receptor binding kinetics.

Techniques

We draw from the methods of mathematical statistics, computer science and engineering. From the characteristics of existing and anticipated data sets and the nature of the research questions, we explore appropriate statistical methods from an inferential and computational viewpoint. Maximum likelihood methods are frequently used to find optimal parameter sets given particular models. Bayes techniques are also investigated where feasible. Graphical and exploratory statistical methods are used to form initial impressions of datasets, and also to communicate findings to scientific colleagues. Simulation methods are useful to characterize the properties of an analysis technique or model. For many models, the effective number of degrees of freedom is a useful measure of model complexity or dimensionality. Penalized or regularized maximum likelihood methods are applicable in this work, especially when considering complex, over-parameterized models. Kernel density estimation techniques are used successfully to predict protein secondary structure. Gibbs sampling strategies are being investigated as promising tools for optimal formulation of several difficult problems. The computational intensity of these methods has required use of powerful workstations and may require workstation clusters for implementation. Careful attention has been paid to the computational complexity of these implementations.

Protein Structure Prediction

It has recently been realized that aligned homologous sequences are useful in predicting an unknown secondary structure from a new protein sequence. That is, having other sequences available, even of unknown structure, is helpful in predicting the structure of a particular sequence, provided those other sequences bear significant overall homology. We have investigated this somewhat surprising result, which adds an improvement of 3 or 4 percentage points to previous success rates. We have evaluated two different prediction schemes on a dataset of proteins whose structure has been recently determined, and it appears that this improvement is quite robust, being more or less equal for both methods and several variations, provided there are more than about 5 sequences available, and the pairwise identity levels are between 20 and 80%.



Schematic drawing of a secondary structure prediction problem. A short, local segment of a protein sequence is used to determine which of three structural possibilities has the highest probability of occurring.

We have also attempted to replicate the high prediction rates attained by another group, using our own quadratic logistic prediction methodology. However, our method consistently performs at about 3% lower rate than the hierarchical neural net model of Rost and Sander. This disparity may be due to subtle

differences in the means by which prediction rates are calculated using cross-validation, or by underlying differences between the neural network and our quadratic-logistic prediction methods. Performance of our method and other methods remain comparable, however, when tested prospectively on a group of twenty proteins not used in the calibration data sets.

The "atlas" of protein structures, prepared by visually comparing the contact maps of each protein, was refined and enlarged. While only a portion of the protein data bank (pdb) can be easily assigned to the alpha, the beta or the alpha/beta classes, about 90% can be thus classified if proteins are broken into "sequence segments." We are refining the definition of such segments in an attempt to make the classification completely objective.

Protein structure prediction remains one of the most difficult and crucial problems in modern biology. Use of the rapidly growing DNA sequence and protein sequence databases to aid in structure prediction is currently one of the most promising areas in this field. Improved statistical methodology will allow more efficient use of such databases, and will be crucial to the success of structure prediction methods. Even modest improvements in secondary structure prediction accuracy may be useful in several tertiary structure or protein folding simulations currently being investigated by others. Critical statistical thinking is needed in evaluating proposals from the many groups working in this area, such as with carefully controlled cross-validation trials of new prediction methods.

We will continue to analyze the protein data bank of protein structures for new patterns and regularities that can be exploited for prediction purposes. For example, we will investigate whether the contact map of a protein, which can be predicted partially from the coordination numbers of its residues, can be better predicted using artificial neural networks. We will investigate further the underlying reasons for the improvement of prediction accuracy using aligned, homologous sequences, and possibly design an optimal approach for utilizing this information. A new description of protein backbone angles will be used in prediction and classification schemes. This description method has the potential to facilitate recognition of subtle folding motifs, which can be incorporated in a generalized secondary structure prediction scheme.

Analysis of Human Growth

Currently, the available methodology for analysis of frequently measured human growth data has several deficiencies, especially in the context of testing for a "saltatory/stasis" pattern of growth. These deficiencies have prevented a maximally efficient test of the "saltatory" hypothesis. Our previous approaches, while sufficient to reject the saltatory growth hypotheses for a set of data collected by C. Heinrichs, D. R. Counts, G. B. Cutler and J. Baron (NICHD, DEB), did not take the full complexity of the time-series of observations into account. Several attempts have been made to solve this problem, culminating in a new method (Reordered Cumulative Growth Analysis, RCGA), which can graphically and numerically depict the degree to which a data set conforms to the saltatory/stasis hypothesis. Using the same set of data, this method now clearly shows which individuals are following the continuous growth model, and distinguishes those who show some anomalies in their growth patterns. Interestingly, the mathematical/statistical methods used in RCGA have immediate parallels in previous work done to analyze pulsatility in 24-hour endocrine time series.

Statistical methods are essential to the interpretation and understanding of complex biomedical and clinical questions. As illustrated by the saltatory/stasis growth model, answers to such questions often are not provided by traditional statistical methods. New methods may be required. Further, new methods sometimes have a surprising generality and can be applied in completely new contexts. The use of endocrine pulsatility methods to analyze the human growth data for the presence of growth spurts or saltations is a prime example of such an application.

Feedback Inhibition of the Hypothalamic-Pituitary-Adrenal Axis by Placental Corticotropin Releasing Hormone

Does corticotropin releasing hormone (CRH) from the placenta in the third trimester of pregnancy influence the "stress" hormones secreted in the hypothalamic-pituitary-adrenal axis (HPA) of the mother? In a study addressing this question, NICHD researchers collected serial blood samples over two twelve-hour periods and measured CRH, cortisol and adrenocorticotropin (ACTH) levels. We analyzed these time series for pulsatility, circadian rhythms and for auto- and cross-correlation. The analysis revealed that while adrenocorticotropin (ACTH) and cortisol are correlated over the 12 hour sampling periods, CRH does not correlate significantly and further does not show a circadian variation. Thus, the evidence does not support a positive circadian regulatory role for placental CRH on the HPA, nor for a regulatory role of cortisol on placental CRH.

Ligand Binding and Kinetic Data Analysis

Two computer programs developed within this section (LIGAND and ALLFIT) are widely distributed and used. Several hundred copies were distributed during the previous year. The Section provided support to NIH users of these and several other programs. The Macintosh® implementations of those programs were significantly enhanced, now providing a wide selection of graphics, which are easily transferred to other programs, and which can be readily produced in camera ready form. Statistical and mathematical modeling consultation and advice were given to several NIH investigators in areas of ligand binding and kinetic data analysis.

Together with R. Shrager (PSL) and E. Rovati (University of Milan), new programs for modeling of 2-site kinetic experiments were written. These general programs make use of features of the MATLAB® package, and are available for many PC and workstation platforms. These same programs, used in a data simulation mode, make possible the rational design of kinetic experiments in more complex situations involving 2 or more classes of binding sites.

Future Plans

Packaging of new statistical methodologies, putting such methods in a form which can be used by other biomedical researchers, is a valuable practical activity of this Section. In the past we have developed and disseminated methods for multiple sequence alignment, multiple binding site analysis, peak detection in endocrine time series, and analysis of multiple parallel curves. We anticipate bringing new methodology to the broader scientific community at NIH in the areas of new alignment techniques, and new protein structure prediction techniques.

Several long-standing and difficult statistical estimation and inference problems seem to be ideally suited to a new statistical calibration technique, the Gibbs sampler. We will investigate several applications of this new approach to such classical problems.

Molecular Graphics and Simulation Section

Bernard R. Brooks, Ph.D., Head

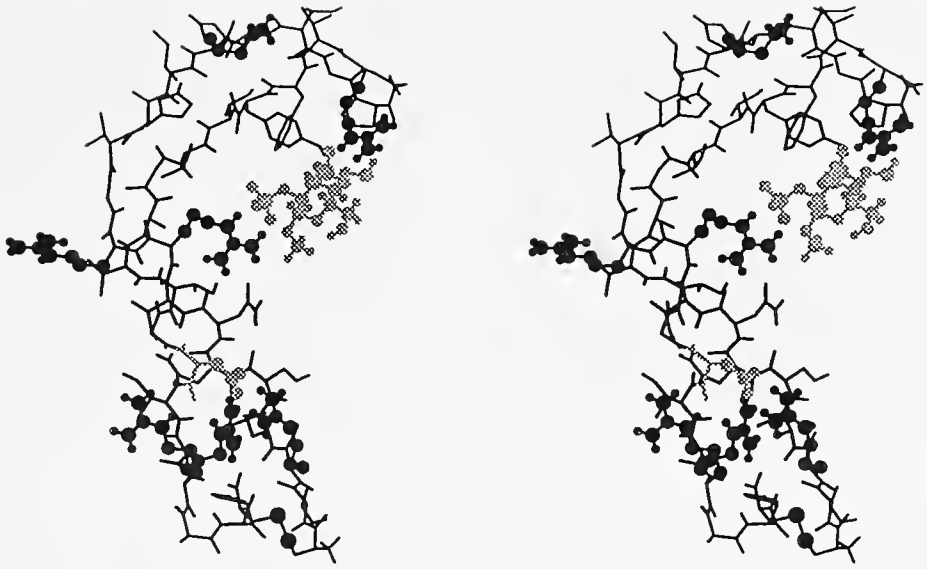
Molecular Dynamics Simulations of Biological Macromolecules

The Molecular Graphics and Simulation Section studies problems of biological significance, using the computational techniques of: molecular dynamics, molecular mechanics, modeling, *ab initio* analysis of small molecule structure, and molecular graphics. These techniques have been applied to a wide variety of macromolecular systems. Specific projects related to the study of HIV proteins include:

- Analysis of inhibitor binding to the active site of HIV-1 protease
- Investigation of the mechanism of action of HIV-1 protease.
- Modeling the V3 loop in HIV-1 gp120 correlating with syncytium formation
- Simulations of HIV-1 reverse transcriptase.

The primary goal of the protease studies is to elucidate the mechanism by which HIV-1 protease binds and cleaves viral polyproteins. The cleavage reaction is a necessary step in the maturation of the HIV-1 virus. Thus HIV-1 protease is a possible target for AIDS therapies, and it is the object of intense theoretical and experimental study. An understanding of the mechanism of reaction would be of great value in the search for effective inhibitors for the protease. Secondary goals of this study include the development of new algorithms to investigate complex reaction processes and the design of inhibitors.

The study of the mechanism of HIV-1 protease was divided into two parts. In the first part, which has been completed, molecular dynamics (MD) simulation was performed, under a variety of starting conditions, to probe conformation space for structures likely to initiate reaction. Proposed mechanisms include general acid-general base catalysis with either a neutral or zwitterion intermediate and direct nucleophilic attack with a covalent intermediate. The simulations suggest that: (1) the catalytic aspartic acid residue (Asp 25) of monomer B is protonated when reaction begins; (2) if the mechanism is general acid-general base, the catalytic Asp of monomer A is protonated when the second reaction step is initiated; (3) the carbonyl oxygen is more likely than the scissile nitrogen to be protonated in the early stages of reaction; (4) the structural water, water 301, stabilizes the transition state but does not participate directly in reaction; and (5) a lytic water, if present, has very little mobility. In the second part of the HIV-1 protease mechanism study, a combined quantum mechanical and molecular mechanical (QM/MM) potential is being used to evaluate reaction barrier heights for those mechanisms whose reactant conformations were shown to be physically accessible in part one. The *ab initio* software packages GAMESS and CHARMM,whose interfacing is described elsewhere in this report, are being used for these calculations. This part of the study entails: (1) determining reaction paths and their energetics for each mechanism by



This is a stereo view of a model of the V3 loop of gp120 of HIV-1 for sequence AMS-165 (syncytium inducing) complexed with a disulfated glucose (mimicking heparan sulfate). The backbone and uncharged residues are shown in line drawing, while the charged residues and the glucose molecule are shown in ball and stick representation. The negatively charged amino acid side residues and the negatively charged glucose are shown shaded.

appropriately searching the potential energy surface; and (2) performing free energy perturbation simulations along the reaction paths in order to provide entropic information. The profile of free energy change along the reaction paths will determine the most likely reaction mechanism.

Model building of the V3 Loop in HIV-1 gp120 has revealed a characteristic positioning of positively charged residues that are correlated with syncytium formation. Structural motifs present in the V3 loops of Syncytium-Inducing (SI) strains of HIV-1, but absent in Nonsyncytium-Inducing Strains (NSI), have been determined. Twenty V3 loop sequences were modeled using CHARMM. Eighteen of the sequences represent serial isolates from nine patients, each of whom underwent a transition from NSI to an SI phenotype. The V3 loop was initially modeled as a beta sheet-type II turn-beta sheet. Energy minimization studies were performed on the hydrated peptide, as was 50 ps of MD simulation at 300 K. The time-averaged structures, averaged over the last 20 picoseconds, were analyzed for their capacity to bind various disulfated sugars. Each of the ten SI strains contains at least two positively charged residues in close proximity, without any intervening negatively charged residues, on one face near the middle of the V3 loop. Each of these V3 loops forms a strong divalent electrostatic interaction with a disulfated sugar. None of the ten NSI strains has two positively charged amino acid residues on one face of the V3 loop. The V3 loop mutated to form the SI phenotype in several ways: alternate amino acids on one beta strand became positively charged, or else an insertion of amino acids brought the arginine at the tip of the loop into close proximity, so it could serve as a positively charged residue.

Other Applied Research on Molecules of Biomedical Interest

In this research, molecular dynamics simulations are used to predict function or structures of peptides and proteins. Such projects include:

- Modeling intermediate filament (IF) proteins
- Identification of peptides that bind to human MHC DR1
- Simulation of a large virus complex, human rhinovirus 14 (HRV14).

Mutations in intermediate filament (IF) proteins are implicated in keratinizing disorders of the skin. The goal of the simulation and modelling of IF proteins is to elucidate the molecular effects of these mutations. A recent report indicates that the 1A domain coiled-coil segments of human keratins 1 and 10 intermediate filaments are hot spots for substitutional mutations, some of which correlate with skin diseases. To investigate the mechanism at the atomic level, MD simulations are being conducted of a wild type and of four mutants of this domain, in a water box with periodic boundaries. Since no crystal structures for this protein are yet available, we model the starting structure as an ideal double helix with known heptad features.

Preliminary results from our simulations have provided some clues to the mechanism of mutation.

Simulations of the virus complex, human rhinovirus 14 (HRV14), have been initiated. In order to obtain a reasonable starting model, and to reduce the computational cost of simulating a system, only 1/60th of the full system is simulated; the simulations probe an asymmetric unit in the full icosahedral symmetry. Once a suitable full atom model of the virus has been simulated with explicit solvation, examinations of the destabilization effects of drug binding will be explored on a 1/12th asymmetric model of the full system. This work is a collaboration with Carol Post at Purdue University.

Basic Research

Basic research is underway to provide a better understanding of macromolecular systems. The projects include studies of:

- Temperature effects on protein dynamics
- The effects of hydration on protein dynamics
- Examining protein anharmonicity, especially the role of dihedral transitions
- Molecular dynamics simulations on staphylococcal nuclease: comparison with NMR data
- Harmonic analysis of large systems
- Modeling and simulation of the lipid bilayers in crystal and gel phases
- Molecular dynamics simulation studies of DNA: the B-Z junction
- The mechanism of lysozyme elucidated by QM/MM techniques
- The mechanism of ribonuclease A elucidated by QM/MM techniques.

The dependence of simulated low-temperature protein dynamics on the cooling procedure has been examined. We have systematically studied the sensitivity of simulated low-temperature protein motion to the method of equilibration. Simulated annealing of dry and hydrated myoglobin was performed at different rates and from different initial temperatures. The discrepancy between low-temperature motional amplitudes, measured experimentally and observed in previous simulations, was reduced by 25%, and efficient equilibration protocols were identified. This work demonstrates that the behavior of proteins at low temperature does depend on how the low temperature 3-dimensional structures are prepared.

Torsional transitions and protein anharmonicity (all motion that cannot be attributed to the harmonicity observed at very low temperatures) have been investigated. Protein dynamics are highly anharmonic at physiological temperatures, and the anharmonic mean-square fluctuation has been correlated with the number of dihedral angles that undergo transitions during simulation. We have investigated the temperature dependence of myoglobin when confined to the torsional substates occupied initially (to prevent crossing of torsional barriers). The simulated dynamics remain remarkably anharmonic, when dihedral transitions are prohibited, and shed light on the role of hydration water as a mobility enhancer. This work demonstrates that solvated proteins do not require dihedral transitions to allow internal motion, but that dihedral transitions are highly correlated with anharmonic mean-squared atomic fluctuation.

Among the topics to be investigated in the future are the hydration and temperature dependencies of the dynamics simulated for non-alpha-helical proteins, the dynamics of myoglobin in solution and crystal environments, and the ongoing optimization of simulation protocols (electrostatic and solvation treatments, among others).

Long time (on the scale of simulation) dynamics of proteins are being studied with simulations of staphylococcal nuclease (SNase). SNase has been widely studied experimentally and provides an excellent system for assessing the accuracy of theoretical methods, and MD provides a way of interpreting experimental measurements. Simulations on the order of 3-4 ns have been performed with SNase solvated by a minimal solvation layer of 350-369 water molecules, for SNase in both its liganded and unliganded forms. The crystal structure of unliganded SNase is maintained to a remarkable degree during simulation, the backbone root mean square deviation (rmsd) remaining less than 1 Å for 3.75 ns while fluctuations are undamped. This provides evidence that the newly developed parameter set used, PARM30, enables very accurate simulations. The motion of SNase has been compared with the experimental measurements of NMR. Simulated backbone motion is in good agreement with NMR measurements; the motion of alanine and leucine side chains will also be assessed. MD can provide insight into the origin of the observed motions. The origin of differences between the structures of unliganded and liganded SNase will also be assessed.

It is nearly two decades since Pohl and Jovin characterized the transition of right handed B-DNA to left handed Z-DNA in solution. However, the nature of the steps involved in transition have remained an unsolved question until now. In order to model the structural transition from B-DNA to Z-DNA, we have developed a new procedure (REPLICA/PATH) in

CHARMM to model transition intermediates by energy minimization and simulated annealing, with two more energy terms connecting replicas: (1) the individual rms deviations between two successive intermediates are close to the average rms deviation; and (2) the path connecting three successive transition intermediates should be nearly linear. A nearly smooth transition has been modeled with 100 transition intermediate structures. These structures are initially obtained as linearly interpolated structures from the initial B-DNA, the final Z-DNA and four proposed intermediate transition models and are refined by minimization with the above mentioned energy terms. The refinement of the transition intermediates is now in progress, using a simulated annealing procedure to improve the quality of the transition pathway.

These projects provide, firstly, insights at the molecular level into complex biological processes and phenomena, which have the potential to influence the design of effective therapies. Secondly, the difficulties and deficiencies encountered in pursuing these projects drives the development of new methods.

Future Plans

In FY95, the MGSS will continue to study relationships between structure and function, and develop the theoretical analysis of inhibitor binding, and specifically study the mechanism of the HIV-1 protease.

Development of Theoretical Methods for Studying Biological Macromolecules

New theoretical techniques being developed are often coupled with software and hardware development. These involve the generation of new simulation techniques and the systematic testing and evaluation of methods. Specific projects include:

- Development of Langevin Piston methods for number/pressure/temperature (NPT) simulation of periodic systems and for stochastic boundary MD simulations
- Development of quantum mechanical potentials and appropriate algorithms for use in molecular dynamics simulations
- Determination of protein structure by NMR and molecular modeling
- Development of an optimized protocol for the preparation of low-temperature states
- Development of flexible MD techniques that remove high frequency degrees of freedom
- Development of the REPLICA/PATH method for determining reaction paths in complex systems using simulated annealing
- Free energy perturbation simulations in solution, examining the effect of restraints
- Conversion of physical models into three-dimensional coordinates for computer analysis and simulation
- Development of Ray Traced Molecular Graphics software for UNIX® workstations, high resolution color printers, and for movies using NTSC video equipment
- Adaptation of a truncated Newton minimizer for CHARMM and biomolecular applications.

A new method, the Langevin Piston (LP), for performing simulations at constant temperature and pressure has been developed and tested. This work extends the seminal work of Andersen, later revised by Nose and Klein. The new method includes a friction term on the piston as well as a stochastic force to achieve a piston that can be described by the Langevin equation. The advantage of this method is that temperature and pressure can be controlled without artificial ringing effects (that occur when no friction is used for the piston), and without sacrificing the ensemble. This method is being extended to allow additional ensembles (e.g., constant surface tension), so that our ability to simulate complex lipid bilayer systems is improved.

One major advance is the combination of the large *ab initio* software package GAMESS and the molecular mechanics program CHARMM. This allows the study of critical portions of a macromolecular system at high accuracy. We are currently working to improve the ability to interface quantum mechanics (QM) and molecular mechanics (MM) regions. In particular, much effort has been aimed at determining how best to handle "link" atoms, the fictitious hydrogen atoms needed by the QM wave function to complete electronic structure shells for closed shell singlet states. These techniques are being applied to numerous model systems before exploring protein mechanisms.

Our research continues to focus, in part, on optimizing protocols for the simulating biomolecules. Among the important issues being addressed are the accurate treatment of solvent effects, the efficient approximation of long-range forces, and the appropriate preparation of low-temperature states.

Analysis of protein dynamics simulation often demands visual representation. For a movie of only few minutes duration, several thousand picture frames need to be generated and stored on video equipment. High quality rendering requires that ray-traced graphic images be generated and stored. Software has been developed that allows these procedures to be performed without

human intervention, so that movies can be made overnight and on weekends and stored on high quality optical disks.

The three-dimensional coordinates from plastic (COORPLAS) program has been enhanced to run with 24-bit planes, using X-windows on HP graphics workstations. A camera is connected to a workstation via a SCSI interface, allowing the user to manipulate images so that three dimensional coordinates for complex models can be obtained from stereo images of plastic physical models.

A truncated Newton minimization method (partial second order method) has been successfully adapted for CHARMM. This new facility improves the performance of large scale energy minimizations. This work was performed in collaboration with Tamar Schlick from the Courant Institute.

Parameters and Models

Parameter sets and models are generally available for most macromolecular systems, but there is considerable room for improvement, and alternate models that improve realism, or reduce computational costs, need to be examined. This effort involves the refinement of parameter sets and the exploration of alternate energetic models for molecules and environmental conditions. Ongoing projects include:

- Evaluation of parameter sets
- Approximation of long-range interactions in macromolecular simulation variants of the Ewald sum method, using a particle mesh grid
- New methods for long range truncation of the energy potential
- Evaluation and comparison of implicit and explicit water models for simulations examining the hydration of proteins
- Molecular dynamics simulation studies of DNA: analysis of the parameter sets using an infinite DNA helix
- Analysis of conformational changes due to solvent: *ab initio* studies.

Evaluating parameter sets for MD simulations of proteins has continued. MD simulations of interleukin hydrated by 350 water molecules were performed under the same protocol for 150 ps by using each of nine parameter sets from Harvard University and Molecular Simulation Inc. This project is a continuation of a previous one, which used carboxy-myoglobin as a sample. Carboxy-myoglobin has typical alpha helical structures, while interleukin has typical beta sheet structures. Analysis based on the last 100 ps of each simulation was done for the rmsd from the starting structure and the rms atomic fluctuations about the average structure. Parameter sets are evaluated based on these results and recommendations are made.

A comprehensive evaluation of spherical cutoff methods used to approximate long-range forces in molecular dynamics simulations, including new methods, was recently published (Steinbach and Brooks, J. Comp. Chem. 15, 667-683 (1994).). These methods improve our ability to perform realistic macromolecular simulations at modest cost.

Development of good methods to simulate macromolecular behavior in a solution is still a problem. Recently, implicit methods using atomic solvation parameters became popular. We have developed software that runs on parallel computers to study the effects of these implicit solvent models. It has been shown that current implicit methods and parameters are inferior to available explicit water models.

Computer simulations of finite length DNA oligonucleotide duplex have been performed earlier by several groups. However, such a simulation always gives rise to end effects and hence may not be a true representation of the natural DNA. We have performed energy minimization, without counterion and water, but with a distance-dependent dielectric, of two DNA polynucleotides of sequence d(A).d(T) and d(G).d(C), by using multiple images of one base pair to simulate an infinite chain. The structures have also been simulated with several different values of twist and rise, to understand the energetics of the twist and rise variability. Although the simulations were performed with the crude approximation of the polyelectrolyte by a simple distance-dependent dielectric, and the findings are therefore preliminary, the simulated structures show a deep minimum energy near the experimentally observed twist (36°) and rise (3.4 Å) of these molecules in B-form. Work is in progress to improve theoretical soundness by incorporating counterions and water molecules and by performing molecular dynamics to overcome local energy barriers.

Future Plans

In FY95, the MGS will continue a broad effort to develop new methods, such as a new integration procedure for MD to improve the accuracy of free energy perturbation simulations. Methods for treating solvent implicitly to provide for hydrophobic effects without the explicit inclusion of many water molecules, and methods to properly treat electronic polarization in molecular dynamics simulations will be explored.

These methods will be applied to a variety of macromolecular systems. Planned projects include:

- Further refinement and examination of free energy techniques
- Development and use of a polarizable and flexible water model
- Three-dimensional structure determination of proteins from a simplified topological description (with Richard Feldmann).

Development of Advanced Computer Hardware and Software

With the increased availability of parallel computer resources amenable to large scale scientific computing, development efforts are essential for optimal use of these resources. The efforts include the development of parallel computing techniques suitable for macromolecular simulation and the development of a parallel workstation cluster for high-efficiency simulations at low cost.

Massively Parallel Computers

Development of methods and software to make productive use of parallel MIMD machines for use in macromolecular simulations is underway. The initial, global communication approach has been successful in providing an efficient full-feature version of CHARMM. This parallel version of CHARMM has been extended to run on almost any MIMD parallel computer platform. Our current development effort involves a scalable algorithm that promises to greatly reduce the communication cost for very large Massively Parallel Processors (MPP) machines or for large workstation clusters. Current projects include:

- A scalable molecular dynamics algorithm for MPP machines and large workstation clusters
- Development and evaluation of parallel algorithms for molecular dynamics
- Development of parallel QM/MM methods
- Development and efficient use of a high speed workstation cluster
- Development of and support for CHARMM.

Our current development effort involves a scalable algorithm that promises to greatly reduce the communication cost for very large MPP machines or for large workstation clusters. The nature of an effective scalable algorithm is that the time spent for communication is reduced as the number of nodes increases. In the algorithm that we are adapting, the communication costs scale as the reciprocal square-root of the number of processors. This is achieved by the "force decomposition method" in which the information that each processor needs to perform its task is limited.

Efficient parallelization of QM/MM methods has proved to be very challenging. The merging of the CHARMM and GAMESS software packages, both of which have been independently optimized to run on MPP machines, provides unique problems in the development of an efficient combined code.

Workstation clusters provide a highly competitive environment in terms of cost performance for macromolecular simulations. A workstation cluster consisting of 17 HP735/755 machines is being developed, as are appropriate hardware and software.

The parallel version of CHARMM developed by this group is being used on many MPP machines, and it has gained widespread acceptance among CHARMM users. This full feature version of CHARMM enables MPP technology to be put to practical use. The parallel version of CHARMM is now being used for most of the research projects in the MGS, and it is proving to be extremely efficient and reliable.

Future Plans

In FY95, we hope to expand the workstation cluster by adding newer HP735/755s servers, and to enhance the communication speed with a high-speed digital switch. A variety of parallel algorithms will be evaluated and put to use on this cluster and on other MPP systems. The new scalable parallel algorithms will be tested and evaluated on a variety of MPP hardware platforms.

Structure and Activity of HIV Reverse Transcriptase

Dhananjay Bhattacharyya with Bernard R. Brooks

HIV-1 reverse transcriptase is the principal target for the treatment of HIV-infected individuals. The structure of reverse transcriptase, with and without complexed nucleic acid, has been solved to a moderate resolution (3 Å) by several groups. However, because crystallographic data was limited, atomic resolution could not be achieved. Thus, only the coordinates of the protein C-alpha atoms and nucleic acid P atoms are in the public domain. We have modeled the complete atom model of the protein/DNA complex (initially solved by Arnold's group at Yale) by a slow buildup procedure using CHARMM. We have also generated a reliable complete atom model of the DNA bound to the protein, using restraint refinement by CHARMM, starting from a B-DNA structure as an initial model, which undertakes a rather unusual structure, with part of the DNA in the B conformation and part in the A conformation. A detailed analysis of this DNA-protein complex, with a few counterions near the solvent

exposed phosphates and a thin hydration shell, has been started, using molecular dynamics simulation to understand energetic stability and fluctuation of the model.

LSB Support Activities

The Molecular Graphics and Simulation Section is actively supporting molecular modeling and simulation needs at the NIH, both through consulting and formal training. Direct services provided by the MGS unit include:

- Research support and guidance for NIH scientists
- Provision of short-term graphics and modeling
- Support for software packages on a variety of hardware platforms
- Examination and evaluation of new hardware
- Assessment of needs at the NIH and policy recommendations to the DCRT management and other DCRT organizations
- Assistance to other DCRT sections in making their computational resources useful for the research needs of NIH.

Courses and Seminar Series

The MGS supports four courses, which are given periodically:

- CHARMM: a Program for Macromolecular Energy, Minimization, and Dynamics
- Usage and Applications of Molecular Quantum Mechanical (QM) Programs
- Molecular Dynamics for Problems in Structural Biology
- Molecular Graphics: Creating Pictures and Videos.

The MGS is also conducting a book review series and an active seminar series for computational chemistry. Both are open to interested scientists.

Future plans

MGS will continue to be a resource for NIH, provide direct collaborative assistance, and give courses and organize seminar series and book review series.

Publications

Bezrukov SM, Kasianowicz. Noise analysis of ionization kinetics in a protein ion channel. In: Handel, PH, Chang AL, eds. Noise in physical systems and 1/f fluctuations. American Institute of Physics: New York, 1993;677-680.

Bezrukov SM, Vodyanoy I. On noise in biological membranes and relevant ionic systems. Advances in Chemistry Series No. 235. Membrane Electrochemistry 1994;375-400.

Bezrukov SM, Vodyanoy I, Parsegian VA. Counting polymers moving through a single ion channel. Nature 1994;370:279-281.

Choe Y, Short BJ, Chen HH, Preisler RS, Rau DC. Osmotic sensitivity of the B-Z transition in DNA. Biophys J 1994;66:A100.

Colombo MF, Rau DC, Parsegian VA. Re-evaluation of chloride's regulation of hemoglobin oxygen uptake: the neglected contribution of protein hydration. Proc Natl Acad Sci USA 1994;91:10517-10520.

Derreumaux P, Zhang G, Schlick T and Brooks BR. A truncated Newton Minimizer adapted for CHARMM and biomolecular applications. J Comp Chem 1994;15:532-555.

Durell SR, Brooks BR and Ben-Naim A. Solvent-induced forces between two hydrophilic groups. J Phys Chem 1994;98:2198-2202.

Heinrichs C, Munson PJ, Counts DR, Cutler Jr GB, Baron J. Patterns of human growth. Science 1994 (in press).

Hodoscek M, Kovacek D, and Maksic ZB. Theoretical study of Mills-Nixon effect in naphto-cyclobutenes and -cyclobutadienes. Theor Chim Acta Berlin, 1993:86:343-351.

Koslov MM, Leikin S, Rand RP. Bending, hydration and interstitial energies quantitatively account for the hexagonal-lamellar-hexagonal reentrant phase transition in dioleoylphosphatidylethanolamine. Biophys J 1994:67:1603-1611.

Leikin S, Parsegian VA. Temperature-induced complementarity as a mechanism for biomolecular assembly. Proteins: Struct Funct Genet 1994;73-76.

Leikin S, Rau DC, Parsegian VA. Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices. Proc Natl Acad Sci USA 1994;91:276-280.

Munson PJ, Baron J, Cutler GB, Heinrichs C, Oerter-Klein K. Do humans grow in fits and starts? A statistical investigation of the properties of the saltation/stasis model. Presentation, DCRT, June 1994.

Munson PJ, Cao L, Di Francesco V, Porrelli R. Semiparametric and kernel density estimation procedures for prediction of protein secondary structure. In: American Statistical Association Annual Meeting,

Statistical Computing Section. Alexandria, VA, 1993;107-113.

Munson PJ, Di Francesco V, Porrelli R. Protein secondary structure prediction using periodic-quadratic-logistic models: Statistical and theoretical issues. In: Hunter L, ed. Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences. IEEE: California, 1994;5:375-384.

Munson PJ, Di Francesco V, Porrelli R. Secondary structure prediction using penalized likelihood models. In: Tartar ME, Lock MD, eds. Proceedings of the 25th Symposium on the Interface on Computing Science and Statistics. California: Statistical Applications of Expanding Computer Capabilities, 1993;203-209.

Parsegian VA. Reconciliation of van der Waals force measurements between phosphatidylcholine bilayers in water and between bilayer-coated mica surfaces. Langmuir 1993;3625-3628.

Parsegian, VA. Strength of van der Waals attraction between mica surfaces across lipid bilayers. Langmuir 1994 (in press).

Parsegian VA, Rand RP, Colombo M, Rau DC. Water at the macromolecular surface: solvation energy and functional control. In: Blank M, Vodyanoy I, eds. Biomembrane Electrochemistry. Washington DC: American Chemical Society, 1994;235:177-196.

Podgornik R, Rau DC, Parsegian VA. Parameterization of direct and soft steric-undulatory forces between DNA double helical polyelectrolytes in solutions of several different anions and cations. Biophys J 1994;66:962-971.

Podgornik R, Svetina S, Zeks B. Parameterization invariance and shape equations of elastic axisymmetric vesicles. Phys Rev E 1994 (in press).

Rau DC, Garner MM. Hydration effects on DNA-protein interactions. Biophys J 1994;66:A370.

Sidorova N, Rau DC. Netropsin binding to specific DNA sequences: coupling of energy and water release. Biophys J 1994;66:A184.

Steinbach PJ, Brooks BR. New spherical-cutoff methods for long-range forces in macromolecular simulation. J Comp Chem 1994;15:667-683.

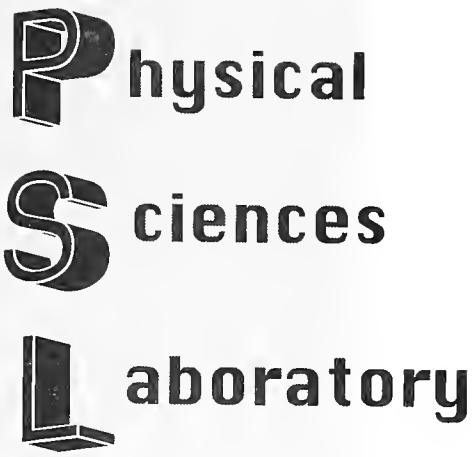
Steinbach PJ, Brooks BR. Protein hydration elucidated by molecular dynamics simulation. Proc Natl Acad Sci USA 1993;90:9135-9139.

Steinbach PJ, Brooks BR. Protein simulation below the glass-transition temperature: dependence on cooling protocol. Chem Phys Let 1994;226:447-452.

Steinbach PJ, Hodoscek M, Brooks BR. Modelling solvation effects on protein dynamics. Biophysics J 1993;64: A183.

Venable RM, Brooks BR and Carson FW. Theoretical studies of relaxation of a monomeric subunit of human immunodeficiency virus type 1 protease in water using molecular dynamics. Proteins: Struct Funct Genet 1993; 15:374-384.

Vodyanoy I, Bezrukov SM, Parsegian VA. Probing alamethicin channels with water-soluble polymers. Size-modulated osmotic action. Biophys J 1993;2097-2105.


Physical
Sciences
Laboratory

Physical Sciences Laboratory

George H. Weiss, Ph.D., Chief

The Physical Sciences Laboratory (PSL) develops methods based on physics and applied mathematics to solve problems of biomedical and biophysical interest. The work of the PSL has two major components: the development of theoretical and experimental techniques for biomedical problems; and consulting and collaboration with NIH scientists on problems for which the methodology of physics or applied mathematics is appropriate.

A major effort by members of the PSL is the use of optical methods to probe biological structure. This aspect of PSL research is directed by Ralph Nossal, and consists of elements of both direct optical imaging of tissues and the utilization of more indirect techniques, such as neutron imaging, to determine properties of biological gels.

New initiatives include a close collaboration with the Nuclear Medicine Department in the Clinical Center. This project develops mathematical and computer algorithms to interpret imaging data.

Biophysical Analysis

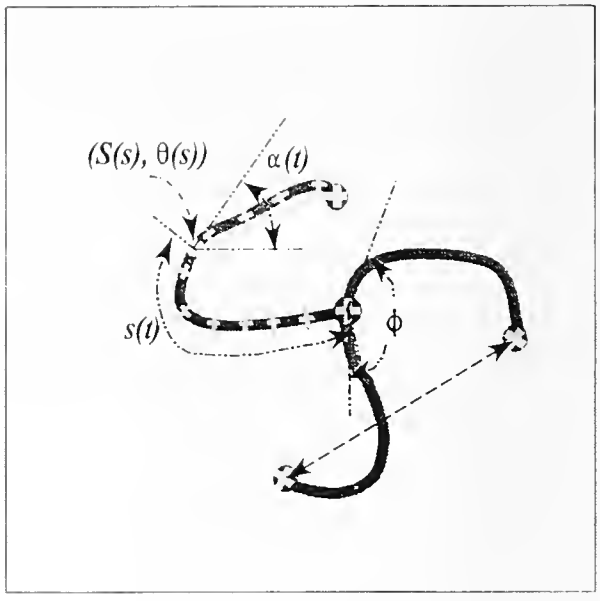
R.J. Nossal, Ph.D.
with A. Gandjbakhche, Ph.D., A. J. Jin, Ph.D., G. H. Weiss, Ph.D. (DCRT/PSL); R. Bonner, Ph.D., J. Schmitt, Ph.D. (NCRR/BEIP); D. Sackett, Ph.D., (NIDDK/LBP); R. Dadmarz, Ph.D., D. J. Schwartzentruber, M.D. (NCI/Surg. Br.); A. C. Steven, Ph. D. (NIAMS/LSB); A. P. Andrews, Ph. D., S. Krueger, Ph.D. (NIST); R. Agah, M.D. (Methodist Hospital, Houston); M. Motamedi, Ph.D. (University of Texas Medical Center); R. Bansil, Ph.D. (Boston University); P. Cerasi, M.S., P. Mills, Ph.D. (Université de Paris).

Quantitative physical and mathematical methods have been applied to several research problems that are of broad interest to biomedical scientists, and can be divided into three categories (cell biophysics, scattering techniques, and tissue optics), which are individually described as follows:

Cell Biophysics

Activities in this area involve using methodologies of theoretical and computational biophysics to investigate phenomena of molecular cell biology. Recent emphasis has been on determining the mechanical and structural properties of large (mesoscopic) molecular structures and on analysis of cytoskeletal elasticity.

Particular attention is being given to networks of clathrin triskelions. The latter are large molecular complexes, which are thought to play a role in the uptake of materials from the extracellular milieu (receptor-mediated endocytosis) by animal cells. We previously studied the topological rules that determine how a network of clathrin triskelions, initially located on a cell surface (a "coated pit"), buds off to form the basket (a "coated vesicle"), which engulfs the material that is being ingested. Focus has been on the lattice rearrangements occurring during the transformation from coated pits to coated vesicles. These involve subtle changes in the structure of constituent triskelions; to gain insight into this aspect of the endocytotic process, we have developed a set of novel analytical and computational tools that relate the shape variations of triskelions to the underlying mechanical properties of the molecules. These new analytical methods are being used to quantitate information obtained from electron micrographs of clathrin triskelions, the goal being to learn about the flexibility of the triskelion arms and the mechanical properties of the central hub where the arms are joined.



Quantitative electron microscopy: systematic analysis of images of a protein supra-molecule, the clathrin triskelion, reveals detailed deformation properties through correlations in shape fluctuations.

Another project relates to the process by which white blood cells (neutrophils) become rigid when acted upon by chemoattractants and other cytokines. Such activated neutrophils may become wedged in the capillaries of the lung, where they can cause significant

pathological distress. Published data on the stiffening of such cells that occurs when neutrophils react with the chemotactic peptide N-formyl-methionyl-leucyl-phenylalanine (FMLP) provide semi-quantitative measures of the development of rigidity in activated cells. We have developed a physico-chemical theory of polymer network formation to understand how the binding of ligands at a cell surface influences the cytoskeletal transformations that give rise to cell rigidity. Presently, the data available in the literature are being analyzed to gain insight into the molecular processes underlying the transduction of the ligand-binding events.

Scattering Techniques

The mesoscopic structure of macromolecular complexes can be probed by diffraction measurements utilizing neutrons or light. During the past year, we have continued our studies of the structure of agarose. This material has long been familiar to biochemists and molecular biologists as a matrix for electrophoretic separation of macromolecules, and it also serves as a model system for polymer matrices that are of interest to cell biologists. Recent emphasis has been on understanding how solution properties affect network junctions, and how gel structure is changed by applied electric fields. Small angle neutron scattering data on agarose gels have been acquired at the National Institute of Standards and Technology research reactor, yielding structural information on length scales of the order of 30 -1200 Å. Results indicate changes in intra-strand associations as hydrogen bonding is disrupted. However, electric field effects are only weakly apparent at such distances; to extend the range of observation, a collaborative small angle light scattering study has been started recently with investigators at Boston University. Preliminary experiments indicate that the gels indeed do distort in electric fields, and experimental apparatus now is being constructed that will allow more careful examination of such effects.

Other light scattering techniques also are being used to examine gel structure. We recently reconstructed a special-purpose dynamic light scattering instrument that enables rapid non-destructive measurement of the elastic shear modulus of a soft gel. This instrument is now being used to study the relationships between intra-strand hydrogen bonding and lattice elasticity in agarose gels and, in collaboration with NIDR scientists, the properties of reconstituted extracellular matrix material ("matrigel" and its derivatives). One extension of this work will be to develop an assay for matrix elasticity when living cells are incorporated in the sample. We then would be able to examine the degradation of extracellular matrix by enzymes released by activated cells, and also assess changes in the mechanical properties of basement membrane materials and other components that modulate cell migration through tissues.

We have also carried out a small-angle neutron scattering ("SANS") study of microtubule assemblies, performed in collaboration with investigators from the Laboratory of Biochemical Pharmacology, NIDDK. The scattering data clearly indicate significant effects of pH and other variables of sample preparation. Changes in microtubule properties are also observed when microtubules are formed in the presence of taxol, an antimitotic drug that is being used in the treatment of breast and ovarian cancer and other neoplastic diseases. We are presently developing analytical expressions to extract quantitative information from the data.

Tissue Optics

The optical properties of tissue play a dominant role in many therapeutic applications of lasers in medicine. Thus, tissue scattering strongly influences the distribution of light within tissues, while the local conversion of photon energy to thermal energy depends on tissue absorption. Because the scattering and absorptive properties of a tissue are modified during therapies that involve photocoagulation or photothermal vaporization, it is important to develop models that allow us to examine how a tissue changes as it is heated. We therefore devised an optically-based, noninvasive method to quantitate thermal damage in tissue. Experiments were carried out to study real-time changes in total reflection and transmission of light due to thermal lesions induced, *in vitro*, in bovine myocardium by a step increase in temperature.

New algorithms, based on a photon random walk treatment of light diffusion, were developed to provide optical coefficients from the measured transmittances and reflectances. These algorithms were used to study the temperature dependent rate processes giving rise to thermal damage. In collaboration with investigators from the National Cancer Institute, we are now using similar methodology to characterize the optical properties of human breast tissue.

Another related project involves theoretical analyses of resolution limits for time-resolved imaging of human breast. Photon migration theory was employed to calculate the Line Spread Function (LSF) of time resolved photons as they cross different planes inside a finite slab. Results were used to determine the spatial resolution for objects imbedded at each depth. The relationships of resolution to scattering cross section,

sample thickness, and photon transit time were determined. It is evident from these calculations that light intensities available at small excess transit times (required to obtain finer resolutions) are strongly dependent on the thickness of the target tissue. These analytical results have been used to examine the feasibility of time-resolved optical techniques for clinical imaging of optically distinct heterogeneities in biological organs. Although we infer vanishingly small detectable photon intensity when such techniques are used to achieve 2 mm resolution in 5 cm-thick tissues, which mimic targets of interest in clinical mammography, we find that submillimeter resolution seems to be achievable with only moderate loss of intensity if thinner targets (e.g., skin lesions) are probed.

Several projects involving tissue optics have been carried out by our group over the past few years. Many of the resulting theoretical and computational developments have been recently summarized in a review, which will soon appear in the journal *Progress in Optics*.

This work shows how random walk theories developed in our laboratory can explain many observations of photon transport in optically turbid media, including several that cannot be treated by conventional diffusion theory. Among these are the quasi-elastic scattering of photons noted in measurements of microcirculatory blood flow, the reflectance of light from layered tissues and media containing randomly-distributed scattering inclusions, and transmittance of light through optically thin slabs.

Instrumentation Analysis

George H. Weiss, Ph.D
with J.D. Bryngelson (DCRT/PSL); S. Pajevic (DCRT/PSL); R. Shrager (DCRT/PSL), S. Sastry (DCRT/PSL); H. Taitelbaum, Ph.D. (DCRT/PSL); S. Bacharach, Ph.D. (CC/NM); R. Carson, Ph.D. (CC/NM); D. Covell, Ph.D. (Frederick Cancer Research Facility); J.A. Ferretti, Ph.D. (NHLBI/IR); M. Garner, Ph.D. (NICHD/LTPB); R. Goans, M.D., Ph.D. (NICHD/LTPB); J. Gruchus, Ph.D. (NHLBI/IR); B. Horwitz, M.D. (NIA/LN); J. van Meter, Ph.D. (NIA); U. Shmueli, Ph.D. (Tel-Aviv University); L. Yaroslavsky (NCRR/BEIP); A. Yergey, Ph.D. (NICHD/LTPB)

An investigation continues into the problem of optimizing NMR magnetization transfer experiments designed to measure physiological rate constants *in vivo*. Such experiments must be carried out quickly because rate constants often change during the time in which measurements are made. An optimized experiment is one in which a maximally precise estimate of the rate constant is made during a specified time. A formalism used for the optimization of NMR measurements of spin-lattice relaxation times has been adapted for the analysis. This has been shown to lead to simple algorithms useful for experimentalists. The project should be completed in the forthcoming year.

A collaboration has been started with members of NICHD/LTPB on the use of ultrasound in the reflection mode for measurements of trabecular bone porosity. If successful, this technique could provide an inexpensive, rapid and noninvasive method for analyzing these properties easily in a general practitioner's office. As such, it could be an important tool for detecting and monitoring osteoporotic diseases. Preliminary measurements have been made mostly on normal individuals, and suggest that bone density decreases continually with age in females; it also decreases with age in males, but seems to level off at around age 65. These measurements also distinguish well between people who run for exercise and those do not. They have also detected an apparently genetically-linked low bone density in some blood relatives of patients with *osteogenesis imperfecta*.

Much remains to be done in developing this technique. The most effective method for processing the ultrasonic images is still to be determined, as well as the level of correlation between ultrasound and standard measurements made with x-rays. Work on this aspect of the general problem has been initiated in collaboration with L. Yaroslavsky of the NCRR. Further theoretical work is required to determine the acoustic properties of fractal media as they affect measured data. Applications of ultrasonic techniques to measure properties of bone in different disease populations will be made in the forthcoming year.

A collaboration has been started with M. Garner of NICHD to determine a resolution parameter to measure the effectiveness of DNA sequencing in gradient gels. Preliminary measurements are being made to determine the dependence of diffusion and velocity of DNA molecules as a function of gel concentration. An earlier mathematical analysis for diffusion in systems with weak diffusion will be applied to the outcome of these measurements.

A problem to be addressed in the next year is a study of the accuracy of protein structures from multi-dimensional NMR and the effect of misassignments of resonance peaks to pairs of amino acid residues. Preliminary estimates of misassignment errors have already been made, and calculations are planned using computationally efficient lattice models of proteins. The theory will be checked in the final stage of the

investigation by programs currently in use at NIH that are not restricted to the lattice structure.

In a new initiative, S. Pajevic and S. Sastry have begun collaborating with the Nuclear Medicine Department of the Clinical Center on various aspects of medical imaging. This project combines aspects of the physics of specific imaging modalities and image processing techniques. S. Sastry is working with J. van Meter (LN/NIA) to eliminate artefacts attributable to the reconstruction algorithm that is applied to positron emission tomography (PET) images. The proposed technique makes use of high resolution images obtained from MRI scans to calibrate the location of different physiological features of the brain. Preliminary tests of the algorithm developed by Sastry have been made on simulated data. The next phase in this investigation involves tests on patient data collected by NIA. The goal of this project is to determine the feasibility of detecting Alzheimer's disease using PET scans.

S. Pajevic is presently working on problems related to the use of dynamic PET techniques to quantitate myocardial blood flow. Computer simulations have been made to determine how both the precision and the bias of estimates of myocardial blood flow are affected by the time shift between the peak found in experimentally obtained curves of arterial and myocardial blood flow. The simulations have been completed, and testing of conclusions will be started on measurements made on patients.

A second major research project is to adapt a currently available simulation program (SIMSET, developed at the University of Washington, Seattle) for studying the motion of photons in a nonhomogeneous medium to the needs of Nuclear Medicine Department of the Clinical Center. This program is designed for the study of PET and single photon emission tomography (SPECT) imaging modalities. At present it is not sufficiently flexible to take all presently available and planned collimator designs into account. Two present subprojects are: (1) development of a more accurate picture of collimator configurations in SPECT; and (2) development of a more user-friendly interface for the program. The first of these will allow photon scattering in SPECT to be studied in greater detail, so that appropriate corrections can be made. When SIMSET has been modified, members of the PSL and of the Nuclear Medicine Department plan to address a number of problems in medical imaging. A major currently unresolved issue in interpreting data from SPECT scans is whether to use all or part of the data collected from SPECT measurements in image reconstruction. A realistic simulation study will be able to settle this and related questions.

A recently started project together with B. Horwitz (LN/NIA) will apply neural network techniques to distinguish normal and schizophrenic brain blood flow patterns using functional brain imaging. This project is in a preliminary phase, the long range goal being to determine a less *ad hoc* technique than those in current use.

A monograph *An Introduction to Crystallographic Statistics* by U. Shmueli and G.H. Weiss has been completed, and will be published by Oxford University Press. It presents the elementary ideas in the use of crystallographic statistics to determine structures from scattering data. It also summarizes research in the past 10 years that replaces approximations by exact, but more computer intensive, formulations of the underlying theory.

Studies in Applied Mathematics and Physics

George H. Weiss, Ph.D.
with S.A. Abrams, M.D. (Baylor College of Medicine, Houston); A.M. Berezhkovskii, Ph.D. (Karpov Institute of Physical Chemistry, Moscow); M. Gitterman, Ph.D., S. Havlin, Ph.D. (Bar-Ilan University); J. Masoliver (University of Barcelona); R. Goans, Ph.D., M.D., A. Szabo, Ph.D. (NIDDK/LCP); P. Wolynes, Ph.D. (University of Illinois); A. Yergey, Ph.D. (NICHD/LTPB).

This project is concerned with the application of mathematical methods to the solution of biomedical and biophysical problems.

Kinetics of Calcium Absorption

A theory of the *kinetics of calcium absorption into bone* has been developed by G. Weiss and collaborators, which extends earlier work on chromatographic models developed in the PSL. The theory is used to interpret measurements of the kinetics of disappearance of labelled calcium from different populations. Preliminary tests of the theory indicate its ability to distinguish between different growth stages in adolescent girls. Work presently being completed shows that the model is also able to detect impaired calcium absorption into bone in patients who are suffering from dermatomyositis and are being treated with steroids. A study of altered calcium absorption kinetics in patients suffering from *osteogenesis imperfecta* is just beginning. Early results indicate that relatives of such patients also show significantly

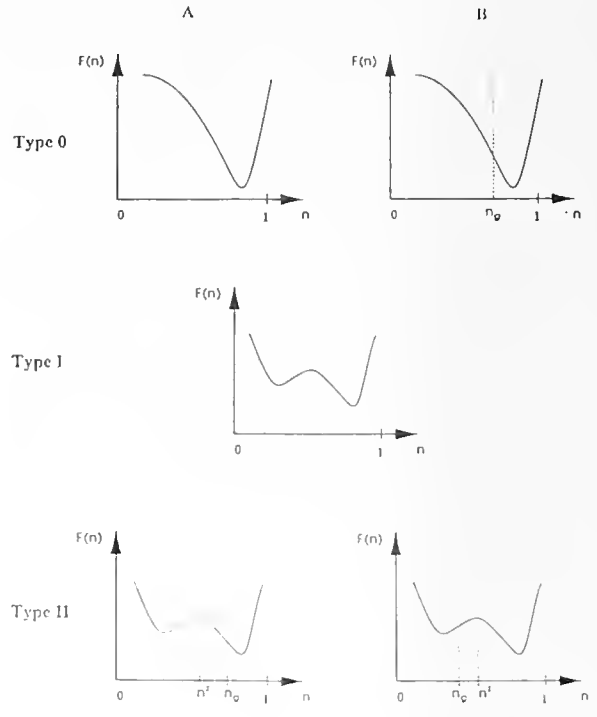
altered calcium absorption, although they have no other symptoms of the disease. This observation will be followed up in the coming year as more subjects become available. Preliminary studies have also been made of patients with glycoside storage disease type I, also showing altered kinetic behavior of calcium.

An associated project is to optimize the protocol for measuring calcium kinetics, which, at present, is both very time-consuming and labor-intensive. The analytical basis for the optimization will be the optimal design theory developed for measuring spin-lattice relaxation times made in NMR experiments.

Protein Folding

The understanding, and even the description of protein folding is impeded by the complexity of the process. J. Bryngelson and his collaborators have developed a theoretical basis for protein folding using the statistical properties of conformational energies. Bryngelson has shown that this statistical approach provides simple, quantitative explanations for: folding as a first-order phase transition; collapsed unfolded states; and the curved Arrhenius plots observed in both these cases and in lattice simulations. Bryngelson has also related these quantitative ideas to folding pathways, the uni-exponential *vs.* multi-exponential behavior in folding experiments and to the effect of mutations on folding.

The theory of Bryngelson and collaborators predicts that a protein can undergo a glass transition while folding. This glass transition occurs when the folding protein's "energy landscape" becomes sufficiently rough, causing the dynamics to slow considerably. The folding scenario observed depends on the conditions under which the folding experiment is carried out. The folding scenarios are illustrated on the figure. In the type 0 scenarios shown at the top, the free energy function has only one minimum near the folded state. In a type 0A transition, shown at left, there is no glass transition. In a type 0B transition, shown at right, at some value of the folding coordinate, n_g, the protein undergoes a glass transition and exhibits the slow, glassy dynamics for the rest of the folding process, that is, for $n>n_g$. In the type I scenario shown in the middle of the figure, the free energy has two minima, an unfolded one and a folded one, and there is no glass transition during the folding process. The free energies in the type II scenarios, shown at the bottom of the figure, also have two minima, but the protein undergoes a glass transition during the folding process. In a type IIA scenario, shown at left, the glass transition occurs after the thermodynamic folding bottleneck, shown at $n^{\ddagger}$ in the sketch. In a type IIB scenario, shown at right, the folding protein is already glassy before the thermodynamic folding bottleneck is reached. Each of these scenarios exhibits different folding behavior. The scenarios where the free energy function has two minima will exhibit nucleation kinetics in folding, whereas the type 0 scenarios will not. Most strikingly, well defined folding pathways will occur in the scenarios where the folding protein undergoes a glass transition.



Schematic illustrations of the protein folding scenarios proposed in the theory of Bryngelson and collaborators. In each sketch, the horizontal axis represents a folding coordinate, n. The larger the value of the folding coordinate, that is, the further to the right in each sketch, the closer the protein molecule is to being fully folded. The vertical axis of each sketch is the average free energy of a protein molecule with that value of the folding coordinate. The folding scenarios are discussed in the text.

Bryngelson and collaborators have studied experimental data on the folding of several proteins and have found that they fit into these theoretical folding scenarios. The theory has already been used to improve the performance of protein structure prediction programs. A related project in progress is to determine the effects of water exclusion in the initial collapse phase in protein folding. Bryngelson has shown that the hydrogen bonds become increasingly effective in determining secondary structure as the protein collapses. This process mimics the simulated annealing algorithm used for complex optimization problems, including

protein structure prediction. When completed, this project should shed light on how real proteins find low-energy conformations, suggest new experiments and new interpretations of old experiments, and may also improve simulation algorithms for protein folding.

Mathematical and Computational Methods for Solving Nonlinear Equations

R.I. Shrager
with G.H. Weiss, Ph.D. (DCRT/PSL); P.J. Munson, Ph.D. (DCRT/LSB); G.E. Rovati, Ph.D. (University of Milan, Italy); M.S. Lewis, Ph.D. (NCRR/BEIP); S-J. Kim, Ph.D. (NCI/DCBDC); S. Bose, Ph.D. (J. Nehru University, New Delhi, India); R. Berger, Ph.D., R. Hendler, Ph.D., (NHLBI/LCB); D.W. Myers, Ph.D., G.K. Ackers, Ph.D. (Washington University School of Medicine, St. Louis); M.L. Doyle, Ph.D. (SmithKline Beecham Pharmaceuticals, King of Prussia, PA); K.D. Vandegriff, Ph.D. (University of California, San Diego, CA); R. Carson, Ph.D. (CC/NMD); U. Shmueli, Ph.D., R. Schach, Ph.D., I. Goldberg, Ph.D. (Tel-Aviv University, Israel)

The purpose of this project is to provide NIH investigators with mathematical tools for insight, analysis, and solution of complex equations arising in the modeling of biological systems. To facilitate these efforts, PSL develops mathematical methods accessible to investigators from many disciplines. Software packages based on these methods are made available to the research community as general research tools. Advice on the use of certain commercial mathematical software packages is also offered:

- Analysis of kinetic binding (with P.J. Munson, DCRT/LSB; G.E. Rovati, University of Milan, Italy): A program for non-linear least squares fitting of binding data is being written, and a paper describing its advantages is in preparation.
- Hemoglobin-oxygen binding (with K. Vandegriff, University of California, San Diego, CA): Studies involving singular value decomposition and multi-wavelength spectrophotometry have been published. The work includes a detailed discussion of the practical differences between rapid-scan spectra and instantaneous (e.g. diode array) spectra.
- Weighted fitting of hemoglobin-oxygen equilibrium curves (with M.L. Doyle, SmithKline Beecham; D.W. Myers, G.K. Ackers, Washington University School of Medicine, St. Louis, MO): Several weighting schemes have been extensively tested by simulation. This work is now published.
- Ultracentrifuge studies of protein-nucleic acid interactions (with M.S. Lewis, NCRR/BEIP, and S-J. Kim, NCI/DCBDC): The use of pseudo-inverses greatly simplifies the analysis of interactions of the type A+B<—>C. A program was written to use the simpler analysis, and a paper showing the advantages of this method has been published.
- Imaging regional cerebral blood flow (with R. Carson, CC/Nuclear Medicine Department): A method for computing cerebral blood flow without an explicit (and invasive) measure of arterial flow has been developed and tested.
- Rapid computation of the probability density function (PDF) for the three phase invariant used in direct methods of phase determination in x-ray crystallography (with U. Shmueli, Tel-Aviv University, Israel) and G.H. Weiss: A program has been written to perform this calculation. The method has the advantage that it is reliable beyond the range of data used to derive it. A paper comparing the method to other rapid calculations is in press.
- Statistics of background radiation from observed x-ray diffraction profiles (with R. Schach, U. Shmueli, I. Goldberg, Tel-Aviv University, Israel; and G.H. Weiss): Crystal structure determination is improved by a better estimate of background radiation. Some phenomenologic estimates have been compared, and a paper is in preparation.
- Bacteriorhodopsin (BR) (with R. Hendler, Salil Bose, NHLBI/LCB): The relaxation kinetics of BR after actinic laser flash is influenced by the intensity of the flash. A paper comparing several models of this dependence is in preparation.
- Matrix tutorial (with R. Hendler, NHLBI/LCB): A tutorial for biochemists unfamiliar with matrices has been published showing the use of singular value decomposition (SVD) and the pseudoinverse in analyzing time-course data from spectrophotometers.
- Floating point arithmetic: Because of the adoption of the IEEE floating point standard as the default arithmetic on CONVEX vector supercomputers, a seminar, including extensive technical notes, was given to the HELIX staff (CFB) to acquaint them with the properties of floating point numbers in general, and IEEE floating point numbers in particular.

Publications

Agah R, Gandjbakhche AH, Motamedi M, Nossal R, Bonner RF. Dynamics of temperature dependent optical properties of tissue. Correlation with thermal damage. IEEE Trans Biomed Engr 1994 (in press).

Borick SS, Debenedetti PG, Sastry S. A lattice model of network-forming fluids with orientation-dependent bonding: Equilibrium stability, and implications for the phase behavior of supercooled water. J Phys Chem 1994 (in press).

Bryngelson JD. When is a potential accurate enough for structure prediction? Theory and application to a random heteropolymer model of protein folding. J Chem Phys 1994;100:6038-6045.

Bryngelson JD, Camacho CJ, Sun Chan H, Dill KA, Onuchic JN, Socci ND, Thirumalai D, Wolynes PG. Kinetics and thermodynamics of protein folding. Nature 1994 (in press).

Bryngelson JD, Onuchic JN, Socci NC, Wolynes PG. Funnels, pathways and the energy landscape of protein folding: a synthesis. Proteins: Structure, Function & Genetics 1994 (in press).

Chakravarthi AS, Debenedetti PG, Sastry S, Yeo S. Thermal expansion and stability limits of generalized van der Waals fluids. J Phys Chem 1994;98:6876-6884.

Dayan I, Havlin S, Weiss GH. Laser beam spreading in transmission through a slab. Lasers in the Life Sciences 1993;5:277-283.

den Hollander F, Weiss GH. Aspects of trapping in transport processes. In: Weiss, GH, ed. Some Recent Problems in Statistical Physics, Philadelphia: SIAM 1994;147-203.

Douek P, Gandjbakhche AH, Leon MB, Bonner RF. Functional properties of a new rheolytic catheter for percutaneous thrombectomy: In vitro investigation. J Invest Radiol 1994;29:547-552.

Doyle ML, Myers DW, Ackers GK, Shrager RI. Weighted nonlinear regression analysis of oxygen isotherms. In: Everse J, Winslow RM, Vandegriff KD, eds., Methods in Enzymology: Hemoglobins, vol. 232, part C, Biophysical Methods. Orlando: Academic Press 1994:576-597.

Eisenberg E, Havlin SH, Weiss GH. Fluctuations of the probability density of diffusing particles for different realizations of a random medium. Phys Rev Lett. 1994;72:2827-2830.

Fisher ME, Jin AJ, Parry AO. On the stiffness of an interface near a wall. Berliner Bunsengesellschaft Phys Chem 1994 (in press).

Gandjbakhche AH, Mills P, Snabre P. A light-scattering technique for the study of orientation and deformation of red blood cells in a concentrated suspension. Appl Opt 1994;33:1070-1078.

Gandjbakhche AH, Mills P, Snabre P, Dufaux J. Aggregation and deformation of red blood cells as probed by a laser light scattering technique in a concentrated suspension: I-Methodology. SPIE 1994;2136:87-96.

Gandjbakhche AH, Nossal R, Bonner RF. Resolution limits for optical transillumination of abnormalities deeply embedded in tissues. Med Phys J 1994;21:185-191.

Gandjbakhche AH, Nossal R, Bonner RF. Theoretical study of resolution limits for time-resolved imaging of human breast. In: Alfano R, ed. Advances in Laser and Light Spectroscopy to Diagnose Cancer and other Diseases. SPIE, 1994;176-185.

Gandjbakhche AH, Othmane A, Mills P, Snabre P, Dufaux J. Aggregation and deformation of red blood cells as probed by a laser light scattering technique in a concentrated suspension: II-Comparison between normal and pathological sets. SPIE 1994;2136:97-104.

Gandjbakhche AH, Weiss GH. Random walk and diffusion-like models of photon migration in turbid media. Progr Optics 1994 (in press).

Gitterman M, Weiss GH. A comment on early-time solution of the Smoluchowski equation. J Stat Phys 1994;74:947-951.

Gitterman M, Weiss GH. A comparison of two methods for solving transport equations with weak diffusion. Sep Sci & Tech 1994;29:1-10.

Gitterman M, Weiss GH. A singular perturbation theory for reaction-diffusion equations. Chem Phys 1994;180:319-328.

Gitterman M, Weiss GH. A transition in a noisy linear system driven by a periodic signal. J Stat Phys 1994;74:941-946.

Gitterman M, Weiss GH. Generalized theory of the kinetics of tracers in biological systems. Bull Math Biol 1994;56:171-186.

Gitterman M, Weiss GH. Exact expansion of moments of the Smoluchowski equation. Phys Rev E 1994 (in press).

Goans RE, Weiss GH, Abrams SA, Perez MD, Yergey AL. Calcium tracer kinetics show decreased irreversible flow to bone in glucocorticoid treated patients. Calcified Tissue Intl 1994 (in press).

Hemric ME, Lu F, Shrager R, Carey J, Chalovich JM. Reversal of caldesmon binding to myosin with calcium-calmodulin or by phosphorylating caldesmon. J Biol Chem 1994;268:15305-15311.

Hendler R, Bose S, Shrager R. Multiwavelength analysis of the kinetics of reduction to cytochrome aa3 by cytochrome c. Biophys J 1993;65:1307-1317.

Hendler RW, Dancshazy Z, Bose S, Shrager RI, Tokaji Z. Influence of excitation energy on the bacteriorhodopsin photocycle. Biophys J 1994 (in press).

Hendler RW, Shrager RI. Deconvolutions based on singular value decomposition and the pseudoinverse: a guide for beginners. J Biochem & Biophys Meth 1994;28:1-33.

Jin AJ, Fisher ME. Stiffness instability in short-range critical wetting. Phys Rev B 1993 (in press).

Jin AJ, Nossal R. Energetic properties of clathrin triskelions inferred from electron microscopic observations. Biophys J 1994 (in press).

Jin AJ, Nossal R. Quantitative estimates of the deformability of clathrin triskelions: analysis of electron micrographs. J Struct Biol 1994 (in press).

Jin AJ, Nossal R. Topological mechanisms involved in the formation of clathrin-coated vesicles. Biophys J 1993;65:1523-1537.

Krueger S, Andrews AP, Nossal R. Small angle neutron scattering studies of structural characteristics of agarose gels. Biophys Chem 1994 (in press).

Larralde H, Weiss GH. The expected number of distinct sites visited by N biased random walkers in one dimension. Physica Acta 1994;209:361-368.

Lewis MS, Shrager RI, Kim S-J. Analysis of protein-nucleic acid and protein-protein interactions using multi-wavelength scans from the XL-A analytical ultracentrifuge. In: Schuster TM, Laue TM, eds. Modern Analytical Ultracentrifugation. Boston: Birkhauser 1994:94-119.

Masoliver J, Porrà JM, Weiss GH. Solution to the telegrapher's equation in the presence of reflecting and partly reflecting boundaries. Phys Rev E 1993;48:939-944.

Masoliver J, Robinson A, Weiss GH. Coherent stochastic resonance. Phys Rev E 1994 (in press).

Masoliver J, Weiss GH. Telegrapher's equations with variable propagation speeds. Phys Rev E 1994;49:3852-3854.

Sastry S, Sciortino F, Stanley HE, Debenedetti PG, Borick SS. Limits of stability of metastable water: Results from lattice models of water. Proceedings of the 12th International Conference on the Properties of Water and Steam, 1994 (in press).

Sastry S, Stanley HE. Low frequency depolarized Raman spectra in water: Results from normal mode analysis. J Chem Phys 1994;100:5361-5366.

Sastry S, Stanley HE, Sciortino F. Low frequency Raman spectra in water by normal mode analysis. In: Bellissent-Funel M-C, Dore JC, eds. Hydrogen Bond Networks. Dordrecht: Kluwer Academic Publishers 1994 (in press).

Schach R, Shmueli U, Goldberg I. Some statistics of background radiation from observed diffraction profiles. Acta Cryst A 1994 (in press).

Schmitt JM, Knuttel A, Gandjbakhche AH, Bonner RF. Optical characterization of dense tissue using low-coherence interferometry. SPIE 1993;1889:197-210.

Sciortino F, Sastry S. Sound propagation in hydrogen bonded molecular liquids: The case of liquid water. In: Bellissent-Funnel, M-C, Dore JC, eds, Hydrogen Bond Networks. Dordrecht: Kluwer Academic Publishers 1994 (in press).

Sciortino F, Sastry S. Sound propagation in liquid water: The puzzle continues. J Chem Phys 1994;100:3881-3893.

Shmueli U, Weiss GH. Effects of non-crystallographic symmetry of the Σ_1 relationship. I. Bicentric arrangement on PI. Acta Cryst 1994 (in press).

Shmueli U, Weiss GH. Introduction to Crystallographic Statistics. Oxford: Oxford Univ Press, 1994 (in press).

Shrager RI. Modelling chemical reaction: the Jacobian paradigm and related issues. Meth Enzymol 1994;240:181-98.

Shrager RI, Shmueli U, Weiss GH. Exact conditional distribution of a three-phase invariant in space group PI. IV. Further improvements of Cochran-like approximations. Acta Cryst 1994 (in press).

Spencer R, Ferretti JA, Weiss GH. Spillover and

incomplete saturation in kinetic measurements. J Magn Res B 1993;101:294-296.

Stosic B, Sastry S, Kostic D, Milosevic S, Stanley HE. Geometric criteria of phase transitions: The Ising model with nearest and next-nearest neighbor interactions. Phys Rev A 1994 (in press).

Taitelbaum H. Segregation in reaction-diffusion systems. Phys A 1993;200:155-164.

Taitelbaum H, Ferretti JA, Spencer RGS, Weiss GH. Optimization of two-stage measurements of T_1. J Magn Res 1994;109:166-171.

Taitelbaum H, Spencer RGS, Weiss GH. Optimization of two-stage measurements of T_1. J Magn Res 1994;109:166-71.

Taitelbaum H, Weiss GH. Anomalous segregation at a single trap in disordered chains. Phys Rev E 1994 (in press).

Taitelbaum H, Weiss GH, Spencer RGS. Optimization of magnetization transfer experiments for kinetic rate measurements. NMR in Biomed 1994;7:287-292.

Vandegriff KD, Shrager RI. Evaluation of oxygen binding to hemoglobin by rapid-scanning spectrophotometry and singular value decomposition. In: Everse J, Winslow RM, Vandegriff KD, eds, Methods in Enzymology: Hemoglobin, vol 232, part C: Biophysical Methods, Orlando: Academic Press 1994:460-486.

Vilensky B, Havlin S, Taitelbaum H, Weiss GH. Momentum effects in reaction-diffusion equations. J Phys Chem 1994;98:7325-7328.

Weiss GH. A primer of random walkology. In: Havlin S, Bunde A, eds. Fractals in Science. Berlin: Springer-Verlag, 1994:2:118-160.

Weiss GH. Aspects and applications of the Random Walk. Amsterdam:North-Holland Press, 1994;361.

Weiss GH. Biographical of Elliott Waters Montroll. Biographical Memoirs 63, Natl Acad Sci 1994;365-380.

Weiss GH. Contemporary problems in statistical physics. Philadelphia: SIAM, 1994 (in press).

Weiss GH. Finite-velocity diffusion. Amer J Phys 1994 (in press).

Weiss GH, Dishon M, Schmitz AM, Bendler JT, Jones AA, Inglefield PT, Bandis A. NMR relaxation in polymers using the Kohlrausch-Williams Watts (KWW) decay function. Polymer 1994;35:1880-1883.

Weiss GH, Gandjbakhche AH, Masoliver J. Isotropization length for random walk models of photon migration in turbid media. J Mod Opt 1994 (in press).

Weiss GH, Gitterman M. An exact expansion of moments of the Smoluchowski equation. Phys Rev E 1994 (in press).

Weiss GH, Gitterman M, Vieira NA, Goans RE, Abrams SA, Yergey AL. A non-Markovian model for calcium kinetics in the body. Pharmacokinetics and Biopharmaceutics 1994 (in press).

Computational

Molecular

Biology

Section

Computational Molecular Biology Section

Peter C. FitzGerald, Ph.D.,
with Robert A. Pearlstein, Ph.D., and Brian Moldover, Ph.D.

CMBS is charged with supporting and guiding the NIH intramural research community in its area of expertise. Section chief Peter FitzGerald has developed a number of new courses, training manuals and booklets for the users of the GCG sequence analysis package and several genetic databases. He has interacted with dozens of scientists from virtually all the ICDs, providing them with assistance and consultation in their analysis of gene sequences. Robert Pearlstein has developed a number of new courses and training materials to assist scientists throughout the NIH with molecular modeling.

The scientific areas addressed by the Computational Molecular Biology Section (CMBS) include the application of computational tools for the collection, analysis and management of primary DNA and protein sequence data, as well as the use of computational techniques to model the chemical structures of proteins, nucleic acids, and other biomolecules. By fulfilling a variety of roles, this group aims to narrow the gap between the very different worlds of "bench research" and scientific computing. Activities include direct support in the application of computational methods, management of central software and databases, user education and training, and the development of novel software tools for the analysis of data and the dissemination of knowledge. Through these activities CMBS interacts with NIH scientists from all ICDs, including individuals from both the main NIH Campus in Bethesda and individuals from NIH satellite facilities.

Computational Molecular Biology

The Genetics Computer Group (GCG®) sequence analysis package, running on the Helix system, continues to be the primary vehicle through which CMBS provides support for computational molecular biology. This past year saw a number of significant changes in the Helix system, which in turn impacted the GCG® software. In May, a hybrid computing environment was formed between a new four processor Silicon Graphics Challenge system and the existing Convex C3830 computer. The new Silicon Graphics system assumed the name, "helix.nih.gov" and the role of primary timesharing system for scientific applications. The Convex C3830 now plays a more dedicated role for computationally intensive applications. This change in hardware had two significant impacts for users of the GCG® software: the introduction of the SGI Challenge marked the final demise of the Covue (Convex VAX/VMS user environment) version of the GCG® sequence analysis software; just as significantly, the Challenge's UNIX® environment corresponds to one of the native development environments for the GCG software, thus ensuring more timely and reliable updates to this software. In its new guise, GCG continues to be as popular as ever with an active user population of more than 400 NIH scientists.

In support of these scientists, who are involved in DNA and protein sequence analysis, CMBS sponsored and ran a number of training courses and seminars to assist them in developing the necessary skills and knowledge to make productive use of the GCG software.

Over the past year the NIH GOPHER Server has been readily accepted by both NIH intramural and extramural scientists as a convenient tool for searching and retrieving DNA and protein sequence data from centrally maintained databases. As an extension of this capability, plans are in place to present a wide variety of molecular biology analysis tools and data through the use of the World Wide Web and MOSAIC (see below).

Publications

FitzGerald PC, Hartley RW. Polyethenoadenosine phosphate as a fluorogenic substrate for barnase. Anal Biochem 1993;214:544-547.

Raghavan N, Freedman DO, FitzGerald PC, Unnasch TR, Ottesen EA, Nutman TB. Cloning and characterization of a potentially protective chitinase-like recombinant antigen from *Wuchereria bancrofti*. Infect Immun 1994;62:1901-1908.

Molecular Modeling and Computational Chemistry

During the past year CMBS has continued to support the NIH intramural community in the area of computational chemistry and molecular modeling. As well as playing a pivotal role in maintaining and managing a variety of software resources (QUANTA®, SYBYL®, Insight II®, MacImdad®), many of them accessible via the NIH network, CMBS has provided guidance, training and expert assistance to numerous NIH investigators.

CMBS training courses in this area included:

- Introduction to Molecular Modeling
- Introduction to SYBYL®
- Introduction to Molecular Graphics
- Introduction to QUANTA® (presented by Don Kyle).

As an active participant in the NIH-wide Structural Biology Interest Group, Pearlstein organized and chaired the session on 5-HT_2 (serotonin) receptors (February 1994) with guest speakers H. Weinstein and J. Shih.

To facilitate the dissemination of information concerning molecular modeling and computational chemistry, CMBS personnel played a key role in setting up the NIH MOSAIC/Web server, and Pearlstein has developed the NIH Molecular Modeling Home Page and the NIH Guide to Molecular Modeling. The latter represents an extension of the work that began as "The SCRC Handbook of Molecular Modeling," an extensive guide to computational chemistry software for NIH scientists. Accessible worldwide across the Internet, via MOSAIC and other network-based browsing tools, the NIH Molecular Modeling Home Page has been recognized by many as the definitive resource on the Internet for information relating to this field. Consisting of an extensive compilation of documentation about the theory and practice of molecular modeling and computational chemistry, this resource also includes specific documentation regarding the facilities and resources available at NIH. Information about the capabilities, strengths and weaknesses of various software products and their requisite hardware platforms is available from this resource. A notable resource accessible from the NIH Molecular Modeling Home Page is the "Molecules R Us" interface to the Brookhaven Protein Data Bank (PDB) (see below).

CMBS established a NIH-wide site-license for MacImdad®, a Macintosh®-based molecular graphics program, in fall 1992. Based on many of the suggestions and recommendations provided by CMBS and other NIH users, a major upgrade to this software was developed and distributed to NIH scientists in the fall of 1993.

Pearlstein has collaborated on a number of scientific research projects throughout this past year. Examples of such projects are: structure-activity relationship studies of novel antitumor compounds; and molecular modeling of the luteinizing hormone (LH) receptor.

Publications

Bivin D.B, Kubota S, Pearlstein R, Morales MF. On how a myosin tryptophan may be perturbed. Proc Natl Acad Sci USA 1993;90:6791-6795.

Lee YS, Pearlstein R, Kador PF. Molecular modeling studies of aldose reductase inhibitors. J Med Chem 1994;37:787-792.

NIH GOPHER and WWW/MOSAIC Servers

The GOPHER Project, jointly run by HPSCS (High Performance Scientific Computing Section of CFB and CMBS, has continued to enjoy much popularity with both the NIH and Internet communities. Making use of GOPHER (University of Minnesota) and WAIS® (Thinking Machines Inc.) software, this client-server based information search and retrieval system provides the Internet community at large easy access to a wide variety of NIH health, grant and programmatic information. In addition, this facility provides NIH intramural scientists with ready access to scientific literature, as well as scientific and administrative data.

As a logical extension of the work on GOPHER, this group introduced the NIH WWW/MOSAIC server in December 1993. MOSAIC is an Internet-based hypertext browser, developed by the Software Design Group (SDG), National Center for Supercomputing Applications (NCSA). This software allows one to discover, retrieve, and display documents and data from all over the Internet. MOSAIC has been developed as part of the World Wide Web project, a distributed hypermedia environment originated at CERN and developed collaboratively by a large informal international design and development team. Like GOPHER before it, the WWW/MOSAIC information system is currently undergoing an explosive growth in use and popularity. With only little of its potential yet realized, the WWW/MOSAIC information system offers great promise for the future.

Accessible from the NIH WWW/MOSAIC Server is the "Molecules R Us" interface to the Brookhaven Protein Data Bank (PDB). This utility was developed by CMBS to provide an easy, user friendly method of accessing the DNA and protein structural data contained within the PDB maintained by Brookhaven National Labs. The "Molecules R Us" interface provides the ability to retrieve an entry as plain text, a custom-drawn image in one of four styles (stick figure, ball and stick, spacefillng or ribbon), and provides an interactive representation with the appropriate client-side viewer program. "Molecules R Us" was presented as part of the WWW94 Chemistry Workshop at the First International Conference on the World-Wide-Web, CERN, Geneva, Switzerland.

The expansion of GOPHER and MOSAIC services is expected to continue into the coming year, with some major enhancements already in development. (See also the report of HPSCS/CFB)

Publications

Casher O, Doherty DC, FitzGerald PC, Goodman J, Hargreaves M, Labanowski JK, Murray-Rust P, Rzepa HS, Winter MJ, Whitaker BJ. Chemistry and the World-Wide-Web. Proceedings of the WWW94 Conference, CERN, Geneva, Switzerland, May 1994 (in press).

Unit of
Electrocardiography
and
Signal
Processing

Unit on Electrocardiography and Signal Processing

J.J. Bailey, M.D., M.Sc., E.W. Pottala, Ph.D., and H.A. Fredrickson
with J. Moak M.D. (Children's National Medical Center, Washington, D.C.), R. Fletcher, M.D. (Veterans Administration Medical Center, Washington, D.C.); K.S. Iyer (NIMH/CNB); D. McAreavey, M.D. (NHLBI/CB); E.C. Phoebus (University of Puerto Rico); K.L.R. Rasmussen, Ph.D. and M. Champoux, Ph.D. (NICHD/LCE), R.W. Bowser, B.Sc. (Creighton University Cardiac Center, Omaha, NE).

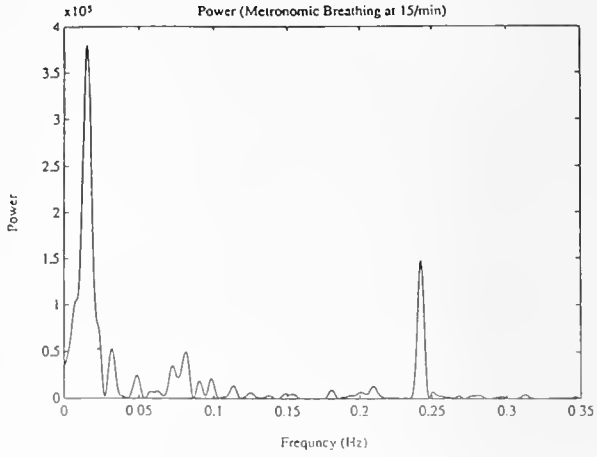
James Bailey has a long-standing interest in the computerized ECG. In recent years, he has been particularly interested in the ambulatory ECG (AECG). When the Laboratory of Applied Science was closed in FY93, Bailey joined the Office of Computational Biosciences. He continues his work in collaboration with NHLBI, NICHD and other medical institutions in the Greater Washington Area and throughout the country.

Computer-Aided Analysis of Electrocardiography

Appropriate use of digital signal processing and automated analysis involving the application of statistically based techniques of information theory and mathematically based engineering methods, together with knowledge of clinical relevance can improve the diagnostic accuracy and prognostic power of electrocardiography, the most widely-used diagnostic tool in cardiology.

These studies currently focus upon the analysis of monitoring electrocardiograms, which are recorded for longer periods (hours to days) than the usual routine ECG (< 30 seconds). ECGs can be recorded in several contexts: i) the 24 hour ambulatory ECG (AECG); ii) laboratory testing using AECG recorders to determine the pathophysiological mechanisms of syncope; and iii) ECG monitoring in intensive care. Many changes in cardiac status can occur within the longer times. Though a single routine ECG may be useful in revealing relatively stable conditions (e.g. hypertrophy), it is a statistically insignificant sample of such changes occurring over hours.

To carry out studies in the first two contexts, we use a SpaceLabs FT2000A Medical Analysis and Review Station (workstation), which is compatible with AECG cassette tapes generated by the Scientific Dynacord Model 423 ambulatory ECG recorder used by the NHLBI Clinical Cardiology Branch, and with cassettes from a number of different model recorders, including Del Mar Avionics, Marquette, and Cardiodata, all used by the Veterans' Administration Medical Center in Washington, DC. Despite extensive literature showing that information extracted by computer analysis of AECGs can be related to cardiac risk factors, there are no standard methods for analysis of AECGs in this rapidly evolving field.



Laboratory spectrogram. Integrity of control of the heart by the autonomic nervous system is tested by spectral analysis of heart rate variability (HVR). In this example the patient was asked to control his breathing following a metronome at 15 times per minute (i.e., frequency of 0.25 Hz). Power spectra analysis of beats detected on electrocardiogram data reveals a peak HRV response at 0.25 Hz, indicating an intact parasympathetic control.

To carry out studies in the third context, we use data from above sources to develop and test real-time methods for suppressing noise (both low frequency baseline wander and high frequency muscle noise), without significantly altering important diagnostic features that could be essential to real-time ECG monitoring in the intensive care unit.

Whether AECG is used in a clinical or research context, the outcome is critically dependent upon the quality and completeness of the data. An important objective of this research is to carry forward previous work in biosignal analysis, with the goal of implementing as much automation as possible to enable and expedite the interpretation of the huge streams of AECG data.

FY94 Progress

The SpaceLabs FT2000A workstation is useful for digitizing the analog ECGs recorded onto cassettes and for previewing and selecting interesting segments of data to be transferred to a host computer with MATLAB™ software. Currently, data is transferred to a Macintosh® system via diskette, which can hold at most about 100 minutes of ECG data (about 1.4

megabytes @ 120 samples/second). Using 100 minute segments of data from AECGs recorded from patients in the VA Congestive Heart Failure study, it was possible to cluster beats into normal and ventricular event categories and to extract statistical parameters from these clusters. A pilot study of several patients is under way.

A cardiology laboratory at the Children's National Medical Center (CNMC) is studying the patho-physiological mechanisms of syncope (e.g., vaso-depression, cardiac inhibition). In these studies, AECG recordings are made while the patient undergoes table-tilt manipulations with and without drugs. The lab's commercial system uses a straightforward Fast Fourier Transform (FFT) of the heart rate variability (HRV) function to produce power spectra, which are unsuitable for distinguishing sympathetic from parasympathetic control of heart rate variability.

In order to produce interpretable power spectra, we modified the techniques in several ways:

- We used RR intervals, which are more primary data than "instantaneous" heart rates, which are derived.
- The Burg algorithm for autoregressive modeling removes the spiky character of the spectrogram resulting from direct application of the FFT with windowing.
- Much of the very low frequency, purported in the literature to relate to renin/angiotensin changes or thermoregulatory influences, overlaps with the low frequency relating to the sympathetic-mediated changes. Therefore, filtering out data under 0.03 Hz allows the low frequency sympathetic peaks and the higher frequency parasympathetic peaks to emerge and be more readily interpreted.
- Instead of absolute values, power density is computed, which is necessary for spectra of successive ECG segments to be comparable.

Pediatric cases tested at the Children's National Medical Center have been analyzed; those subjects whose symptoms were reproduced by table-tilt and abolished by drug therapy have shown typical changes in the power spectra. Other subjects in whom testing did not reproduce their symptoms did not show these typical changes.

Spectra from a few patients with cardiomyopathy who had undergone table-tilt and metronomic breathing manipulations at the NHLBI Clinical Cardiology Branch had been previously collected on cassette and stored. These spectra are now being re-analyzed with the new techniques.

Collaborative Projects

Several collaborative projects were undertaken that involve acquiring data and developing analysis techniques for physiologic signals (e.g., electrocardiogram, electroencephalogram, blood pressure) or for signals obtained from laboratory apparatus (e.g., mass spectrometer, gas chromatograph).

General tasks may involve the development of methods to analyze very large data sets (e.g., 12 hours and 16 channels of electroencephalographic data). These methods include data reduction, noise suppression, pattern recognition, and statistically or mathematically based feature extraction, trend analysis, and construction of spectra, where appropriate. Visual inspection of the power-density spectrogram or trend graphs of extracted parameters often reveal the essential information. However, in some contexts (e.g., mass spectrometer) the data may consist of a spectrum with overlapping peaks and the objective would be to resolve its principal components.

During FY94 a data acquisition board from National Instruments was obtained for the Macintosh® for use in the real-time acquisition of analog wave forms. This approach uses the NI-DAQ software drivers written in C. Using this package to obtain the data, the analysis software can be adapted with MATLAB™ to the particular need of the laboratory. For example, the Clinical Neuroendocrinology Branch, NIMH, needs to acquire and process 16 simultaneous channels of analog data. The LCE Branch, NICHD, needs a real time system that can control stimulus/response experiments in animal studies. A digital I/O board has been ordered for this study.

Development of signal processing techniques for the analysis of ECG data has concentrated on:

- Accurate R-wave detection
- The use of auto-regressive modeling techniques to generate spectra of heart rate variability
- Classification and clustering of beats based on a set of comb filter values for the QRS-T complex
- Real time suppression of baseline wander and high frequency muscle noise.

These improved signal processing techniques, with knowledge of clinical relevance, can improve the diagnostic accuracy and prognostic value of the ECG. The detection and spectral analysis techniques have been implemented and are currently used to correlate changes in autonomic control of heart rate with behavior in monkeys (NICHD/LCE).

Future Plans

Connection of the Spacelabs workstation to the Convex and SUN® workstations via the network will be a major step forward in this project; it will allow 24 hours' worth of data to be transferred to the SUN® workstation. Processing on the SUN® workstation will be at least 30 times faster than on the Macintosh® and will allow analysis of larger groups of patients within reasonable times. The pilot studies of cases from the Children's National Medical Center and those from the Veterans' Administration Congestive Heart Failure project will be used to design protocols for larger patient series.

Continued support will be given to the development of software for the RRV monkey studies performed by the LCE, which includes spectral techniques to analyze the RRV data. The monkey stimulus response system should be operational soon, and the data analysis begun. The general data acquisition system will also be available for routine use in the near future, and the 16 channel EEG study is under way.

Publications

Bailey JJ, McAreavey D, Pottala EW. Methods for testing autonomic control of heart rate: table-tilt and metronomic breathing manipulations. J Electrocardiol 1993;26(suppl):34-36.

Bailey JJ, Pottala EW, Rasmussen KLR. Techniques for enhancement of RR interval power spectra in short epochs of simian monitoring ECGs. Computers in Cardiology, 1994 (in press).

Laks M, Arzbaecher R, Bailey JJ, Berson A, Briller S, Geselowitz D. Will relaxing safe current limits for electromedical equipment increase hazards to patients? Circulation, 1994; 89:909-910.

Moody GB, Feldman CL, Bailey JJ. Standards and applicable databases for long term ECG monitoring. J Electrocardiol 1993;26(suppl):151-155.

Pottala EW, Bailey JJ, Gilham J. Effect of timing resolution upon RRV spectra with a robust QRS detector after band pass filtering. J Electrocardiol 1993;26(suppl):38.

Pottala EW, Rasmussen KLR, Bailey JJ. Effect of missing data on RR interval variability power spectra in simian ECGs: a simple method for filling gaps. J Electrocardiol 1994;27 (in press).

Office of Computing Resources and Services

Office of Computing Resources and Services

J. Emmett Ward
Associate Director

Looking back on its first full year of operation, the Office of Computing Resources and Services (OCRS) can identify a number of satisfying accomplishments.

We consider *TASC*, the DCRT Technical Assistance and Support Center, to be one of DCRT's most important initiatives in our effort to serve the needs of the NIH computing community. During FY94, various helpline and service functions within the Division have been slowly assimilated under the aegis of the TASC in order that "one call" to 4-DCRT can "do all". The tracking system that the Customer Services Branch (CSB) uses to support this effort is a valuable resource to account for, follow-up and resolve questions, issues and problems. It also represents a potentially valuable storehouse of the resolutions of trouble calls and the answers to information requests. Eventually, we hope to tap this knowledge-base and to use it as a prospective facility for new inquiries, and for identifying opportunities to improve and add services such as "just-in-time" documentation and direct LAN support.

The OCRS has strengthened and expanded DCRT's contacts and two-way communications with the NIH community through a variety of vehicles, such as the Architectural Management Staff, which fosters technical and management participation in the process of establishing standards for computing and networking at NIH; helpline consulting services and electronic mail; working partnerships with the ICDs; sponsorship of and/or active participation in Campus special interest groups like the Campus Users Resource Exchange (CURE), Biomedical Research Macintosh Users Group (BRMUG), Wordprocessing Users Group (WUG), and the Office Technology Coordinators (OTCs); and support of the Lead User concept at NIH.

To increase the efficiency and effectiveness of DCRT support, this year the Distributed Systems Branch (DSB) combined the Lead User and the Macintosh® Support Coordinator programs into a single Computer Support Coordinator program. Training was sponsored by DSB for computer support coordinators in PC hardware troubleshooting and in Microsoft® network operating systems. In response to multiple ICD requests, DSB agreed to initiate support for Novell® Netware. DSB also conducted a study of energy-efficient computers during 1994, which provided guidance and recommendations for implementing the Executive Order to purchase Energy Star micro-computers. The results were distributed via e-mail and PCBriefs.

While the OCRS continues support of one of DCRT's most frequently accessed mainframe systems, the Statistical Analysis System (SAS), FY94 saw increased support of SAS on desktop workstations. This was expedited by increasing the number of SAS products on the PC from 5 to 13 and by negotiating DCRT-subsidized site licenses for campus users of PCs, SUN® SPARC workstations, and Macintosh® microcomputers.

During FY94 the Network Systems Branch (NSB) extended the NIHnet to connect the majority of buildings on the campus, including the Children's Inn, and also the NIH facilities at Executive Boulevard, Parklawn, Federal Building, and the SURAnet regional office in College Park, MD, which is NIH's Internet access point.

Increased use of the Internet by NIH components has been facilitated by the Computing Facilities Branch (CFB) jointly managed support and operation of the NIH GOPHER (a client-server based information search and retrieval system) and the operation of a World Wide Web (WWW) information server. During the year, an expanded base of NIH information was made available to intramural scientists and to the international Internet community by the addition of CANCERLIT, NIH Consensus Conferences Statements and NIH Technology conferences and Workshops statements, and the NIH Job Vacancies list to an already useful set of online databases.

The CFB introduced a long-awaited capability to the NIH community during FY94, the Enterprise Print Service. This facility represents a major breakthrough for users of mainframe systems who wish to direct output to their local printers. For users of Delpro and other ADB services, DB2, Wylbur, and batch jobs, output can now be directed to printers at their local PCs, workstations, and LAN servers.

Major improvements to CFB's DB2 database management system environment help to enhance its use for emerging client/server applications. The CFB has announced support for both the Sybase and the Oracle gateways, providing client access to DB2 from user-owned workstations.

As part of its shift from a research and development activity to a full production service, the Advanced

Laboratory Workstation (ALW) system began a three-year transition to full cost recovery. During FY94, one-third of the costs of providing ALW services were recovered from user fees, while two-thirds were subsidized by the Management Fund. Concomitantly, ALW use grew by 33%, with the system now supporting more than 200 client workstations and over 400 users.

A new UNIX®, six-processor Silicon Graphics (SGI) Challenge system was introduced to scientific users in May. This system will offload general purpose tasks from the NIH Convex system. While the user environment closely resembles that of the Convex, the SGI system has been configured to handle such tasks as editing, e-mail, file transfers, and modest computations. Computationally-intensive scientific tasks requiring large memory and vector processing continue to run on the Convex system.

The ISB's successful implementation of two LAN-based client/server solutions to the strategic and business goals of the Clinical Center's Medical Record Department was completed during FY94. Although the Microfilm Index system and the Chargeout System are two separate systems, they are merged and presented to the user as a single system using relational database and client/server technologies. Within two months of operation, the CC noticed improved efficiency in the operation of its various organizational entities. This pilot application of client/server technologies in the NIH environment demonstrates the feasibility of using a multi-vendor, LAN-based solution to developing systems that provide a Graphical User Interface (GUI) front end to a SQL database.

As we begin the transition to 21st-century computing, DCRT is fostering a major ongoing initiative, a NIH-wide cooperative effort to establish a technology architecture. The goal of this project is to achieve consensus on standards for a technical architecture that will meet NIH and ICD scientific and business requirements for computing and networking. Choosing a consistent technical architecture can help NIH to substantially reduce training and support costs, reduce the burden of tracking technology and upgrades, improve communications among the various platforms at NIH, and cut time and effort for file and format interchange.

A major acquisition, called the Computer Equipment, Resources and Technology Acquisition for NIH, Project CERTAN, was begun during the past year to address NIH intramural and extramural support, and administrative information technology requirements. The objective of this project is to provide computing and communication equipment, software, and services to enable DCRT to provide scientific and technological support to the NIH ICDs into the 21st century."

A Taste of the Future

The ISB is planning to improve the user interfaces to the Administrative Data Base (ADB) and the ADB Information System (ADBIS). Today's systems use 3270 screen technology. Recent studies estimate user productivity improvements of about 37% when organizations change from 3270 technology to GUIs at the desktop. GUIs will be developed as front-ends to the ADB and ADBIS to provide a more intuitive environment to each application user and to establish a bridge to client/server computing.

Closely related to the DCRT's overall strategic plan is the challenge to re-engineer the ADB and the Central Accounting System (CAS) software. In re-engineering, one captures the functionality of the current system through a reverse engineering process and then forward engineers the system applying new technologies. Through this approach, DCRT can leverage the investment that has been made in these systems over the years. DCRT will then be positioned to migrate these systems to new technologies when they are available.

Over the next 12 to 24 months, the Scientific Computing Resource Center (SCRC), DSB intends to upgrade its software resources and to deploy new computational tools. It also plans to add a combination demonstration and work area to accommodate small groups of scientists, and to provide network access to SCRC software from anywhere on the campus. A scientific and technical graphics facility is also being considered.

In line with its strategy of ensuring that NIH is state-of-the-art, the NSB will prototype an Asynchronous Transfer Mode (ATM) network during FY95. Within the next two years, CFB will introduce technologies similar to those used in the ALW system to support PC and Macintosh® microcomputers, as well as technologies to provide central backup and archiving of LAN data. The results of Project CERTAN will also accrue over the next few years. In early FY96, we hope to have a support services contract in place, and by fiscal year end, we anticipate the award of the enterprise system contract.

By deploying a standards-based Fiber Distributed Data Interface (FDDI) backbone now and tracking the ATM technology, the NSB is meeting today's needs

while positioning the NIHnet backbone for tomorrow's technologies. The NSB is also working to provide dial-up access to the NIHnet to enable NIH personnel at home or while travelling to connect from a PC or Macintosh®. This will be one more step in the DCRT's efforts to move NIH to a truly distributed computing environment.

Statistical Support Staff

Ray Danner, Chief

The Statistical Support Staff (SSS), which is part of the Office of Computing Resources and Services, comprises five individuals with mathematical, statistical and programming backgrounds. It is responsible for providing NIH scientists and administrators with a wide range of services concerned with the application of computer technology essential to NIH programs. This group provides:

- selection, maintenance and support of standard mathematical/statistical software for general use of research investigators in the NIH community ; support includes training, advice and assistance on proper use of the available software
- collaboration and service in a wide range of computational aspects of biomedical data analysis
- advice and consultation on the quantitative analysis of biomedical research data and use of the computer in such analysis.

SSS provides statistical, mathematical, and other scientific systems and packages to the NIH user community and evaluates new systems and packages for suitability to NIH needs. Computer systems and packages supported by SSS are shown in the table on page 67.

As in previous years, the SAS statistical and data management system was very extensively used at NIH, with an average of 71,700 accesses per month via the IBM 370. SPSS was accessed around 1,900 times per month, and the BMDP® package was accessed an average of over 300 times per month.

SSS mainframe statistical support included:

- maintenance of the system or package
- adequate documentation of NIH computer system changes, system or package updates, and corrections
- rapid response to queries about user access to the most used systems and packages
- procurement of the software.

The SSS answered over 3,500 calls for software assistance, handling requests for information on job control language (JCL), program parameters, uploading and downloading program and data files, and other operating system procedures, as well as assisting in interpretation of results. SSS continues the support of mainframe statistical systems and documentation in response to NIH computer system changes, product updates, and corrections.

Other mainframe software supported by SSS had more limited use. Support for IMSL has included the Convex as well as IBM 370 mainframes. There were relatively few sessions for such specialized programs as GLIM® and RPART (see table).

While NIH-wide use of statistical software on PC and Macintosh® microcomputers is more difficult to quantify, SSS has continued to expand its support of software on these increasingly popular platforms. During FY94, SSS converted about 50% of the SAS/PC copies from MS-DOS to WINDOWS.® SSS took advantage of SAS®'s Enterprise Computing Offer (ECO) to increase the number of SAS® products on the PC from five to thirteen and increase the number of user copies from 300 to 400. Also under the ECO the SAS® system was licensed for OS/2. SSS also continued support of SAS® on the SUN® SPARC workstation under UNIX.® Three SUN® SPARC workstations were purchased, so that SSS can expand this effort. SSS began supporting LIMDEP® on the IBM 370 and is investigating the possibility of making it available on the PC. A site license was negotiated for StatExact for the PC and should be in place by the end of FY94 or the beginning of FY95.

Recognizing the importance of teaching the effective use of systems and packages to biomedical researchers and other NIH users, SSS maintained a substantial program of short courses, prepared documentation and held informational talks. Enrollment in the SAS courses continued at a high level. SSS taught six SAS courses, a total of 15 times, to over 150 students through the DCRT training unit. SSS contracted to have two statistical courses taught through the NIH Training Center. These two courses were very popular, with 30 students attending. SSS also presented four seminars, which were attended by a total of 110 people.

Future Plans

SSS's high level of support for IBM 370 statistical software systems will be continued. More statistical software will be supported on the SUN® SPARC workstations, DOS- WINDOWS®-based PC, and Macintosh.® SSS will continue to support the MS-DOS, WINDOWS,® and OS/2 SAS® site license on the PC. SSS also plans to procure a site license for JMP on the Macintosh® and offer full support; SSS is

in the initial stages of acquiring a site license for SPSS/PC. SSS will support SPSS and BMDP® on the SUN® SPARC workstation. A statistical, mathematical, and graphics software fair was held during the second week of October 1994, which gave SSS information about the NIH requirements for mathematical and statistical software.

Systems and Packages supported by SSS

SAS, SAS/GRAPH, SAS/ETS, SAS/OR, SAS/FSP, SAS/AF SAS/IML, SAS/CBT101, SAS/CBT102, SAS/CBT106, SAS/INSIGHT, S/QC, AS/CALC, SAS/TOOLKIT, SAS/ASSIST, SAS/DB2, SAS/CONNECT, SAS/STAT
Vendor: SAS Institute, Inc. A batch and interactive IBM 370 system for statistical analysis with extensive file manipulation and graphics capabilities, which is also supported under MS-DOS and WINDOWS.®

RPART
Public domain SAS procedure, which performs the recursive partitioning analysis routines of J. H. Friedman.

LIMDEP
Vendor: Econometric Software, Inc. An IBM 370 batch system for econometric modeling.

BMDP
Vendor: BMDP Statistical Software, Inc. A collection of programs for univariate and multivariate statistical analysis supported on the IBM 370 and MS-DOS and WINDOWS®.

LISREL, PRELIS
Vendor: Scientific Software, Inc. A batch IBM 370 program that estimates the unknown coefficients of a set of linear structural equations.

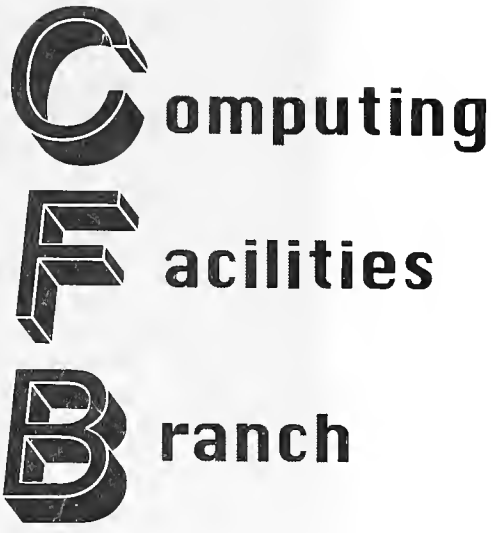
IMSL (International Mathematical and Statistical Libraries)
Vendor: Visual Numerics, Inc. An extensive collection of FORTRAN routines for statistical and mathematical analysis supported for IBM 370, Convex and MS-DOS machines.

SUDANN, SESUDAAN, SURREG, RATIOEST, RTIFREQS, RTILOGIT
Vendor: Research Triangle Institute. Batch and interactive IBM 370 software for sample survey data analysis.

MSTAT1
Source: DCRT staff. IBM 370 batch programs and subroutines for mathematical and statistical analysis.

GLIM (Generalized Linear Interactive Modeling)
Vendor: Numerical Algorithms Group, Inc. An IBM 370 batch and interactive system for analysis of linear statistical models.

SPSS, SPSS-PC+
Vendor: SPSS, Inc. A system for univariate and multivariate statistical analysis with file handling capabilities supported in batch mode on the IBM 370 and interactive mode on IBM 370 and MS-DOS and WINDOWS® machines.


Computing
Facilities
Branch

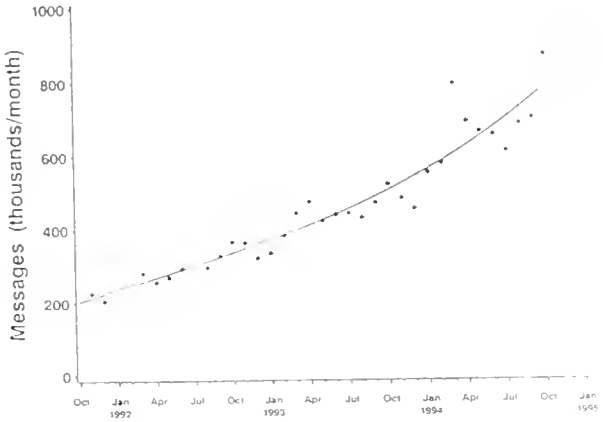
Computing Facilities Branch

Perry S. Plexico, Acting Chief

The Computing Facilities Branch (CFB) plans, implements, maintains, operates, and supports centrally owned or administered computing resources for both the scientific and administrative programs of NIH. The branch also strives to achieve interoperability among the resources it provides and with other computing facilities owned by other organizations in the NIH community. CFB is the outgrowth of the Computer Center Branch, which for 27 years provided computing and networking services to NIH research investigators and administrators who conduct and manage modern biomedical research. In addition, CFB includes important components of the former Computer Systems Laboratory (CSL), with its emphasis on support of computing for the scientific laboratory, and the Advanced Laboratory Workstation (ALW) project.

The NIH Computer Center and its associated telecommunications facilities are among the resources for which CFB is responsible. The Computer Center is made up of interconnected multicomputer facilities designed around large-scale IBM 370 mainframe and Convex supercomputer systems. CFB also has responsibility for DCRT's ALW project, an open-systems approach to distributed computing systems. A centrally administered distributed file system supports UNIX® workstations connected to NIHnet. CFB is building on these technologies to provide full interoperability among the branch's computing resources, and between these resources and user-owned personal computers and workstations.

NIH Internet and Bitnet Traffic



CFB provides interactive timesharing, database, and batch computing services on its mainframes to approximately 17,000 authorized users at NIH and in other agencies throughout the Federal government, on a fee-for-service, full cost-recovery basis. High performance scientific computing services and systems are funded largely by the NIH Management Fund, and are thus restricted to NIH staff. For ALW services, FY94 was the first year of a three-year transition from Management Fund to full cost recovery.

Thousands of NIH users on local area networks (LANs) can access all major facilities (IBM, Convex, and ALW) over NIHnet, the campus-wide area network. Mail gateways allow central facility and LAN users to exchange electronic mail among themselves and with others via the Internet and BITNET. The Federal Telecommunications Service (FTS), international 800 service, and commercial switched telephone lines provide traditional telephone connectivity.

CFB provides its services through the Office of the Chief and six sections:

- The *Office of the Chief* sets branch policy, guides strategic planning, and manages CFB activities by coordinating the work of the sections to encourage and ensure cooperation and integration of efforts. The Office of the Chief also exercises principal responsibility for capacity management, disaster recovery, financial management, procurement and property management.
- The *Database Systems Section* (DBSS), led by William Jones, evaluates, implements, and supports central database systems and tools. Its responsibilities include IMS, DB2, and various database servers and gateways to help users implement client/server access to centrally managed data.
- The *Distributed Systems Section* (DSS), led by Keith Gorlen, investigates and evaluates distributed computing technologies, and applies these to develop, implement, and support networked, interoperable, open-architecture, distributed computing environments. DSS has principal responsibility for the Advanced Laboratory Workstation services.
- The *Enterprise Systems Development Section* (ESDS), led by Elliot Alterman, plans, manages, supports and coordinates hardware and software systems integration and development work related to the current IBM 370 mainframe facility and to other platforms that CFB will introduce in the future for corporate use on an NIH-wide basis.
- The *Enterprise Technologies Section* (ETS), led by Oliver Morton, identifies, evaluates, documents,

and supports products (other than database) to address the software capabilities needed on the current IBM 370 platform and for NIH-wide corporate use with future open-systems architectures. ETS also serves as the principal link between CFB and the Customer Services Branch (CSB), providing second-level user support for enterprise systems.

- The *High Performance Scientific Computing Section* (HPSCS), led by John Dickson, plans, manages, and supports centrally located high performance computers specifically designed for scientific use at NIH, and works toward incorporating them into an NIH distributed computing environment. This section currently supports the Silicon Graphics and Convex systems.
- The *Systems Operations Management Section* (SOMS), led by Robert Mamayek, manages the maintenance and operation of central computing facilities, the physical plant supporting them, and their physical security. It also provides operations and maintenance support for NIHnet and offers plotting and output distribution services. SOMS has principal responsibility for introducing automated operations into the Computer Center, including evaluating, implementing, and managing software tools and robotics.

Highlights of the Year

Last year's major reorganization of the Division of Computer Research and Technology provided the Computing Facilities Branch the opportunity to expand its role in offering computing and data processing resources to the NIH community, while continuing to improve the efficiency and effectiveness of its services. A number of services previously provided by CFB, including the help desk, training, and technical information services, were transferred to the new Customer Service Branch, and CFB took on additional responsibilities in keeping with its new mission.

Major improvements to the Computer Center's DB2 database system environment were made to enhance its use for emerging client/server applications. CFB now supports two gateways (from Sybase and Oracle), which offer client access to DB2 from user-owned client workstations. Several "intelligent" gateways, which offer a more sophisticated client interface to DB2 by providing SQL conversion, were evaluated for possible future support. Other DB2 improvements include: a major rate reduction (up to 80% for some applications); increased availability (from 3:00 AM to midnight daily); increased table sizes; and a broader training curriculum.

A new UNIX® system, a six-processor Silicon Graphics (SGI) Challenge system, was placed in service this year to augment the computational services available to the NIH scientific community. By off-loading general purpose tasks (such as e-mail, routine scalar computing, and GOPHER access) from the Convex system, which was running at full capacity, the new system will improve performance on these general purpose tasks and permit better use of the Convex system for the computationally-intensive applications best suited to it.

The replacement of the IBM 370 system's nearly one terabyte of online disk storage with new technology began this year; the transition is scheduled for completion in the first half of FY95. The new devices offer improvements in reliability, performance, environmental factors and cost effectiveness. By using the System Managed Storage (SMS) component of the MVS operating system, also implemented by CFB this year, the change should be essentially transparent to users. Hierarchical Storage Management (HSM) moves data from one device type to another without involvement by users. Together, SMS and HSM offer further benefits to users, including simplified data set allocation and more flexible dataset backup, migration and recovery.

As a part of its shift from a research and development activity to a full production service, the ALW system began a three-year transition to full cost recovery. This year, one-third of the costs of providing ALW services were recovered from user fees, while two thirds were subsidized by the Management Fund. Concomitantly, ALW usage grew by 33%. The system now supports more than 200 client workstations and over 400 users.

Architectural Management

Architectural management deals with developing and managing an information technology (IT) architecture, a set of standards for computing and networking that apply across an organization and allow the organization to use its information technology assets more effectively. A consistent IT architecture can help NIH optimize its use of IT staff and resources by promoting greater transferability of skills and knowledge, thus potentially reducing FTEs or contract support staff, and by increasing opportunities for bulk purchases. Moreover, an IT architecture promotes interoperability among dissimilar systems and fosters application, data, and platform connectivity among new

and legacy systems. But the most important benefit is the increased productivity that will result when sufficient transparency is achieved that, from their laboratories or desktops, users can access and use the platforms, processes and data required for their work.

Architectural standards are of great interest to CFB, since it offers a wide variety of computing services through a number of computing platforms. Under CFB leadership, the DCRT Architectural Management Staff (AMS) has started the task of identifying candidate standards for computing and networking by involving the user community in efforts to define an architecture and by seeking consensus and commitment on the part of NIH management and IT users.

The Gartner Group was enlisted to assist in developing a draft strategic architecture and related tactical initiatives for the next three to five years. The Gartner Group provided an essential industry perspective, which helped achieve a common understanding of architectural issues and terminology. The Gartner Group facilitated a three-day AMS retreat in October 1993, which explored initial guidelines for an architectural strategy. The retreat resulted in a report, which became a reference for an NIH user retreat.

The subsequent three-day NIH user retreat, also facilitated by the Gartner Group, was convened during the winter of 1994. Approximately 25 members of the NIH community participated, including at least one representative from each ICD. Participants were nominated by their respective ICD Executive Officers. This retreat closely paralleled the DCRT retreat in format, covering the following topics:

- Client/server distributed computing model
- Client platforms
- Server platforms
- Network operating systems
- Physical network infrastructure
- Network protocols for internetwork communications
- Electronic mail
- E-mail directory services
- Computer and network security
- Software asset management, including electronic software distribution and site licensing
- Application development (tools/methods).

From this user retreat, the Gartner Group prepared a report for NIH review. The report will serve as the basis for seeking a broad commitment to the proposed IT architecture by NIH and ICD management. Because of the rapid pace of computing and networking innovation, the IT architecture must evolve over the coming years, requiring continuing ICD and DCRT planning and refinement. For example, areas of architectural importance that will need to be addressed by joint user-DCRT working groups include document management, decision support/executive information systems, data access and management, transaction processing and systems and network management.

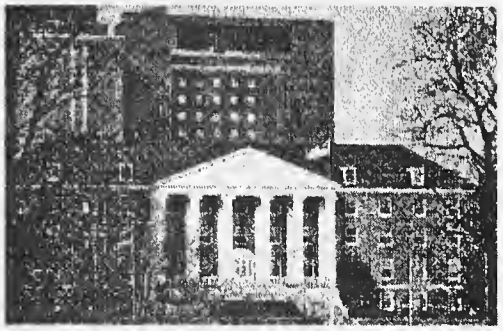
Major Procurement Effort Under Way

A major acquisition called Project CERTAN (Computer Equipment, Resources and Technology Acquisition for NIH) was begun during the past fiscal year to address NIH scientific, extramural support, and administrative information technology (IT) requirements. The objective of this project is to provide computing and communication equipment, software, and services to enable DCRT to provide scientific and technological IT support to the NIH ICDs into the 21st century. Most of the investment in CERTAN will cover central computing resources and services to be managed by CFB.

The assistance of the General Services Administration (GSA) Federal System Integration and Management Center (FEDSIM) was enlisted to develop a Concept of Operations (CONOP) to conduct a requirements analysis for the procurement and to prepare Statements of Work for the contracts necessary to satisfy the requirements. The CONOP identified concepts that provide strategic direction, recommended improvement to the current system, and identified areas in which users want DCRT to provide central computing support.

A Trail Boss acquisition team for Project CERTAN {consisting of staff from DCRT, the Office of Information Resources Management, FEDSIM, the Federal Computer Acquisition Center (FEDCAC) and the Naval Regional Contracting Center (NRCC)} has been established, as has a CERTAN Advisory Council, whose members are drawn from senior level management of NIH, PHS, HHS and GSA. The Trail Boss is a member of the CFB staff, as is a Deputy Trail Boss (Technical), who has responsibility for central mainframe requirements. The Advisory Council will provide policy guidance to the acquisition team, periodically review the status of the acquisition, and address issues presented by the acquisition team.

The acquisition team, with the help of a FEDSIM support contractor, interviewed small groups of users in February and March 1994 to identify information technology requirements. The support contractor used the interview data to develop a formal requirements



National Institutes of Health

Welcome to the National Institutes of Health (NIH) Mosaic Home Page, maintained by the Division of Computer Research and Technology (DCRT).

National Institutes of Health (NIH) Gopher
Search NIH Gopher Menus for topics of interest
NIH Gopher Users Guide

Topics

- Biomedical Information Relating to Health Issues and Clinical Protocols
- NIH Grants and Contracts including the CRISP database
- Research Opportunities at NIH from the NIH Office of Education
- Topics relating to Molecular Biology and Molecular Modeling
- NIH Special Interest Groups
- NIH Computer and Network Resources
- NIH Calendar of Events this week and next week (available after Wednesday)
- Search NIH Phone Listing and Email directory service
- NIH Bethesda Campus Info
- NIH Library On–Line Information Services
- Other NIH Information Services
- Information Superhighway ON–RAMP
- About this Server

(**NOTE:** *This World Wide Web hypertext server is currently under construction and is subject to rapid change. Please send comments to gopher@gopher.nih.gov*)

analysis and is assisting the acquisition team in developing functional specifications for draft statements of work late in FY94. The CFB staff, as well as other members of DCRT staff, have made major investments of time, energy and resources to provide input and review acquisition documents produced by the contractor. Completion of the draft Statements of Work will mark the transition of primary support for the CERTAN procurement from FEDSIM to FEDCAC, and NRCC, which will prepare Requests For Proposals (RFP), and conduct solicitations.

Four procurements are planned, covering enterprise systems, scientific systems, distributed computing resources, and support services. FEDCAC will prepare and release draft RFPs and solicit industry comments for the enterprise, scientific and distributed resources contracts. Appropriate comments will be incorporated into the final RFPs. Source selection plans will identify the detailed criteria on which the vendor's technical proposals will be evaluated. After FEDCAC issues the RFPs, a Source Selection Evaluation Board, with members drawn from NIH and GSA, will

evaluation, negotiate and establish contracts with the winning vendor(s), and issue the first delivery order. Control and responsibility for the contracts will then revert to NIH.

The fourth contract, for support services, will be awarded and administered on behalf of NIH by the NRCC at the Washington Navy Yard, using similar procedures.

Advanced Laboratory Workstations Increase Productivity

More than 400 NIH scientists now use Advanced Laboratory Workstations (ALW). An ALW consists of a high-speed UNIX® workstation, high-resolution color monitor, and optional PostScript printer provided by the user. All other services are provided by CFB, which installs the latest operating system onto the workstation, operates Andrew File System (AFS) file servers, backs up user directories nightly, offers in-house UNIX® training, and provides a variety of state-of-the-art scientific application programs.

Five ALW fileservers were upgraded to newer technology SUN® SPARCserver 1000s, which provide more than 160 Gbytes of storage capacity. The upgraded servers improve performance, reliability and capacity. New versions of applications programs were installed, including two PC favorites: WordPerfect® and Lotus® 1-2-3.

ALW has implemented its own hypertext World Wide Web (WWW) server for disseminating information to its growing user population. ALW users can run the MOSAIC client program to access information on the latest improvements to the ALW System, an extensive list of biomedical institutions, medical databases from around the world and on line documentation. The server presents this information, using a combination of text, graphics, images, and sound.

New General Purpose UNIX System for Scientific Computing

A new UNIX® system, a six-processor Silicon Graphics (SGI) Challenge system with 384 megabytes of memory, was introduced to users in May 1994 to augment the computational services currently available to the NIH scientific community. This system serves primarily to off-load general purpose tasks from the NIH Convex system, which was running at full capacity every weekday 24 hours a day. CFB planned the acquisition to create a central heterogeneous computing environment for NIH scientists, consisting of a general purpose system and computation servers to run applications and programs optimally configured for their hardware capabilities.

Typical general purpose tasks include: editing; reading, and sending electronic mail; transferring files; and modest-sized computations. Analysis of workload data has shown that as many as 90% of the nearly 2,000 Convex users active each month utilize the system for such general purpose computing. The SGI system has been configured to handle these applications optimally with a user environment that very closely resembles that of the Convex. Computationally intensive scientific tasks, requiring large memory or vector processing, continue to run on the Convex system.

Because most current network services are only available on the SGI system, the network name "helix.nih.gov" was reassigned to the new system and the name "coil.nih.gov" was assigned to the Convex system. All users with accounts on the NIH Convex system were given accounts on the SGI. Ultimately, use of the Convex system will be limited to those whose work requires its applications, compilers, and hardware.

Top Ten GCG Programs According to Use

Program Name	usage*
SeqEd	1625
BestFit	1594
Fetch	1047
Fasta	977
Gap	705
Map	462
StringSearch	459
Reformat	416
Translate	395
PileUp	237

* *Usage of GCG programs on the NIH Helix System for the month of September 1994*

This new heterogeneous scientific computing environment provides NIH scientists with improved interactive response for general purpose computing needs on the SGI and enables users with computation-intensive scientific applications to process large jobs on the Convex while receiving the fastest possible turnaround times.

Because the SGI system provides routine computer services comparable to other fee-for-service offerings of CFB, it was determined that this new service should also operate on a cost recovery basis. Although Management Fund monies paid for the initial hardware

and software configuration, continuing operation, support, and upgrades would require a modest user subscription fee, initially estimated at $20 per month, to begin in FY95. The Convex system will continue to be financed by the Management Fund.

Computer Center Software Improves

A number of improvements to software and system management on the IBM 370 system took place this year. The transition from the OS/VS COBOL compiler to COBOL/370 was completed, with the COBOL/370 compiler becoming the production version of COBOL in June.

A new MVS security environment took effect in January 1994, which requires all users to be registered to the Resource Allocation Control Facility (RACF). In this environment, user initials are also RACF identifiers, and RACF passwords must be replaced every six months. Data security on the NIH Convex was also enhanced in January, when all user accounts were modified to require a new password at least every six months.

A new release of the MVS operating system, Data Facility System Managed Storage (DFSMS) was installed during the past year. An enhanced release of IBM's Enterprise System Architecture (ESA), this new level of the operating system offers significant enhancements to the service and reliability of the NIH Computer Center.

NIH Centralized Bulletin Board Achieves Milestone

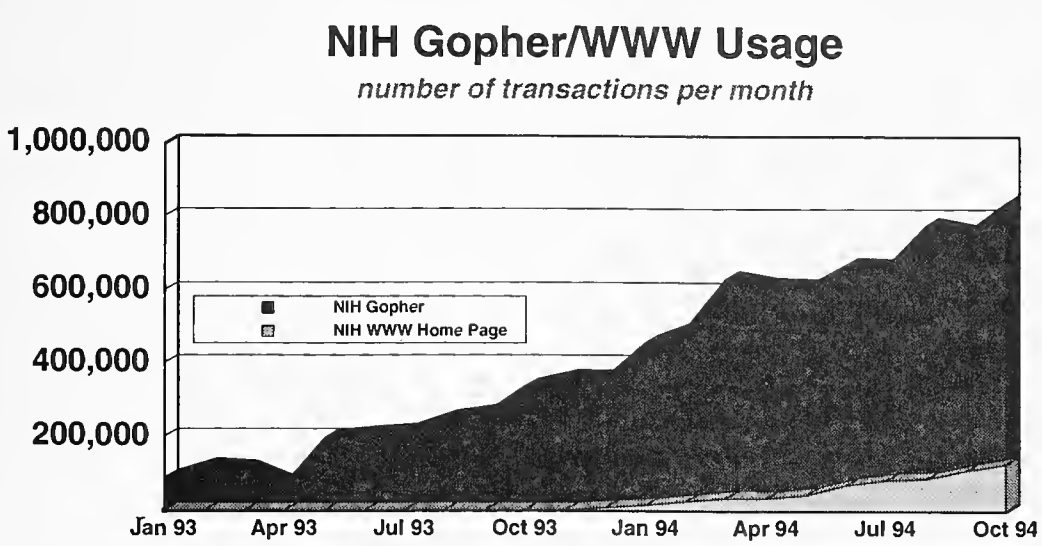
Four years ago the NIH Centralized Bulletin Board System (ENTER BBS) was made available to facilitate creation, administration and access to bulletin boards by the Computer Center's IBM 370 mainframe users. In September 1994, a major milestone of usage was reached with this popular system: the 100,000th bulletin board access was achieved!

GOPHER and MOSAIC Provide Access to NIH Information

CFB and the Computational Molecular Biology Section (CMBS) in DCRT/OD jointly manage and operate the NIH GOPHER, a client-server based information search and retrieval system. GOPHER was developed at the University of Minnesota. During the past year this project has been expanded to include the operation of a WWW information server, accessible via MOSAIC client programs.

These services make a wide variety of NIH information available to NIH intramural scientists and to the international Internet community. Such information includes:

- NIH guidelines on health issues and clinical protocols: CancerNet, CANCERLIT, NIH Consensus Conferences Statements, NIH Technology Conferences and Workshops statements
- NIH grant and contract and programmatic data:



NIH WWW Home Page introduced December 1993 *Accessed by more than 100,000 systems since January 1994*

CRISP (1993 & 1994), NIH Guide to Grants and Contracts, NIH Office of Education Catalog, NIH Inter-Institute Structural Biology Catalog.

NIH scientists have made extensive use of these information systems to gain ready access to an array of scientific data, including:

- Molecular biology databases
- Images of Protein Data Bank (PDB) protein structures
- Weekly updated bibliographic reference data (Current Contents, Reference Update)
- CANCERLINE (Physicians Desk Reference for physicians and patients).

GOPHER and MOSAIC have also been very popular in providing searchable access to such administrative resources as:

- NIH E-mail Directory
- NIH Phone Book
- NIH Job Vacancies.

FY94 also saw GOPHER enter the world of electronic publishing, by providing electronic access to the NIH Deputy Director's DDIR bulletins and "The NIH Catalyst" publication.

With more than 29,000 accesses per day from a total of more than 2,100 different client machines (both from NIH and around the world), GOPHER continues to grow at a rate of approximately 5% per month. In FY94, approximately 500 NIH personnel attended seminars and training classes describing the NIH GOPHER and MOSAIC Servers. (See also the CMBS report).

DB2 Database Environment Improved

A number of improvements were made to the DB2 database environment during the past year. These enhancements made DB2 a more efficient and economical platform as a database server, for both interactive mainframe and client/server access.

The enhancements to DB2 help to make the database system a viable server in the rapidly emerging client/server environment. More than two years of testing and evaluation have been conducted by CFB staff on client software and gateways that can provide state-of-the-art Graphical User Interface (GUI) access from networked workstations to DB2 databases. Demonstrations of these products by the CFB staff and by users who have designed and written client applications have been met with enthusiasm by the user community. Two gateways to DB2 (for Sybase and Oracle) have been installed in the NIH Computer Center as level 2 production services. Both of these gateways allow client applications to send SQL requests to DB2 and present the results in a graphical user-friendly format. They do not, however, translate SQL from Sybase or Oracle to DB2, so that applications must generate DB2-compatible Systems Query Language (SQL). CFB is in the process of evaluating three more capable gateways, which would provide SQL translation and the "look and feel" of the native database system to the client product. Initial tests with a Sybase gateway of this type were successful, and additional tests suggest great potential for these advanced gateways.

A redesign of the DB2 maintenance and backup system made it possible to extend the availability of DB2 for both interactive and batch processing. The new hours are Monday through Saturday from 3 a.m. until midnight and Sunday from 8:30 a.m. until midnight. This extension of DB2 hours allows DB2 database administrators and users greater flexibility in accessing data, and is especially helpful when data needs to be accessed by people who are not in the Eastern time zone.

The maximum size supported for a DB2 table on public DB2 volumes was quadrupled. CFB is actively seeking to offer even larger tables for public DB2 volumes.

Procedures for recovering a DB2 table from a backup were enhanced to enable users to list all available backups for a given table and to generate all required operator tape mount messages.

CFB is investigating the feasibility of providing enhanced database administration services to users. This would include: designing databases; layouts creating tables and indexes; tuning and scheduling database loads when necessary. Such services would allow users to support centrally stored data without having the technical knowledge necessary to perform fundamental DB2 database administration tasks.

Commercial software that significantly reduces the time and cost of loading large DB2 tables has been acquired for the Computer Center. Initial tests have shown that the software may reduce the time of loading some tables by as much as 70-80%. CFB is conducting further testing and evaluation of the product to determine whether it can be made generally available.

System Managed Storage

The move toward a System Managed Storage (SMS) environment for the IBM 370 system made considerable progress during the past year. After

extensive preparation, which included converting the existing Operating System (OS) catalogs to Integrated Catalog Facility catalogs, consolidating public disk storage, and establishing an all-cataloged environment for public disk storage, CFB has taken a series of steps towards implementing the SMS environment. Limited access to SMS began in January 1994, and automatic routing of VSAM (Virtual Storage Access Method) and BDAM (Basic Direct Access Storage Method) data sets to SMS was started on March 1. In April, all data sets destined for public {FILE, MSS (Managed Storage System), TMP (Temporary)} disks began being redirected to SMS. By year's end, essentially all user data had been moved to SMS.

SMS works with the Hierarchical Storage Manager (HSM) to provide effective storage allocation and management. SMS simplifies storage data set placement and backup requirements. It also allocates space to improve disk utilization and meet performance requirements. HSM manages disk volumes and data sets to ensure that active and inactive data are properly placed and that appropriate backups are taken.

Some of the benefits of System Managed Storage are:

- simplification of dataset allocation
- more flexibility in dataset backup, migration, and recovery
- better performance
- cost reduction
- easier transition to new disk architectures.

New Disk Storage Devices

The online Direct Access Storage Device (DASD) environment of the NIH Computer Center, which has consisted of 3380-K devices installed in the 1980s, is in the process of being replaced. Although the 3380-K devices have performed well, reliability has decreased with age, and some capabilities needed for existing and planned services are lacking. CFB is replacing all online DASDs with newer technology devices, Redundant Arrays of Individual Discs (RAIDS). RAIDS will improve reliability and performance, consume less power and require less cooling. This combination of benefits makes the upgrade a cost-effective step to maintain the critical online storage environment.

The approximately one trillion bytes of online storage associated with the IBM 370 system is being upgraded to new technology on a capacity, not a device, basis. This means that total online storage capacity is unchanged, but significantly fewer individual devices will be needed. The resulting reduction in the number of volumes will have no effect on public storage, although fewer dedicated storage devices will be available.

The SMS and HSM components of the MVS operating system are handling the conversion of data from the current public storage environment to the new one in a way that is essentially transparent to users.

Cost Savings Passed on to Users

Rate reductions for most mainframe services were once again passed on to users of the NIH Computer Center, making this the 26th consecutive year of falling prices. Disk storage, batch processing, and interactive services all received significant reductions.

Rates were reduced for most data storage classes. The largest reduction was for FILE volumes, which were reduced 33.3% from $0.0015/track-day to $0.001/track-day. Migrated data and MSS volumes each received a 25% reduction, from $0.001/track-day to $0.00075/track day. The cost of dedicated volumes dropped 10%, from $1,500 per month to $1,350 per month.

The cost of batch processing continued to drop as well. Changes were made to some of the coefficients used in the charging algorithm, which was simplified at the beginning of the last fiscal year. Although it is not possible to determine how any one job benefitted from these reductions, the average savings for batch jobs was more than $375,000 per month, a 30% reduction in overall batch processing costs.

The largest portion of this year's rate reduction went to interactive services: WYLBUR, TSO (including DB2), and IMS. The cost for WYLBUR and TSO services was reduced 35%, from $0.68 to $0.44 per second, and the cost of IMS service dropped 25%, from $0.08 to $0.06 per transaction. The cost for interactive access to DB2 was reduced again in March 1994, when a new charging algorithm for DB2 services was put in effect. The new charging methodology reduced the cost of accessing data stored in DB2 by an additional 65%.

Development of New High Speed Communications Access Continues

CFB staff, assisted by users, tested new COMTEN communications controllers and state-of-the-art modems for higher speed dialup access to the IBM 370 interactive services. As a result, several adjustments were made to configurations of both the communications controllers themselves and to the new modems, so

that they would operate properly in the NIH computing environment. The COMTEN was made available to users for testing in January 1994. This user testing revealed that interactive services respond differently to the "break" key when accessed at higher speeds via the COMTEN than is the case in the current production communications environment. This will require added user assistance by CFB and CSB and some changes to previous operating procedures by users, as the COMTEN controllers are introduced into production service early in FY95.

New Service Provides Local Printing from System/370

To accommodate user requirements, the CFB developed support for local printing of output generated on the IBM 370. Printed output from Delpro and other ADB services, DB2, Wylbur, and batch jobs can now be directed to printers attached to a PC, workstation, or local area network (LAN) connected to NIHnet. This service is provided through the Enterprise Print Services (EPS) product from Interlink Computer Sciences, Inc. EPS also supports Appletalk printers in conjunction with the PrintShare services offered by the Network Systems Branch.

Progress Towards Automated Operations

Progress continued this year toward achieving an automated computer operations environment that will reduce operational costs and improve service to users. An interagency agreement with Office of Personnel Management (OPM) provided access to an automated operations consulting firm, which is helping CFB determine the areas that can best benefit from automation and is assisting with evaluation of software products for some of these areas. These consultant services will be expanded by an 8(a) contract with the same firm, expected to be awarded early in FY95. In addition, planning was completed for new printers equipped with roll paper handlers and for post-processing equipment that will simplify printer operation and output distribution. Preliminary evaluation was done of new tape library management software, which is a necessary prerequisite to the future implementation of automated tape libraries ("silos").

First Disaster Recovery "Hot Site" Test Successful

CFB continued its efforts to establish a disaster recovery plan for the NIH Computer Center. A formal contingency plan was developed, and a test system was made available to individuals responsible for 14 critical applications, which run on the NIH Computer Center mainframes and require recovery within 72 hours following a disaster. Applications were successfully packaged and tested to run at a hot site if a disaster should occur. A new contractual agreement signed early in the fiscal year provides hot site services at a commercial disaster recovery site in New Jersey to satisfy NIH needs in case of a disaster. CFB staff members performed a test at the hot site, during which they restored the NIH operating system and ran a subset of the Test Job Stream. Elements of one of the critical applications also was tested. Future hot site tests, to be conducted approximately twice per year, will encompass the remainder of the 14 designated applications. Once these have been tested successfully, CFB will expand contingency planning to include other applications.

Documentation Services

This year, six editions of INTERFACE, CFB's series of technical notes, were published. Two major updates of the two-volume "Computer Center Users Guide" were issued. Production and dissemination of documentation to users of the NIH IBM and Convex mainframes was transferred to the newly created Customer Services Branch of DCRT during the past year.

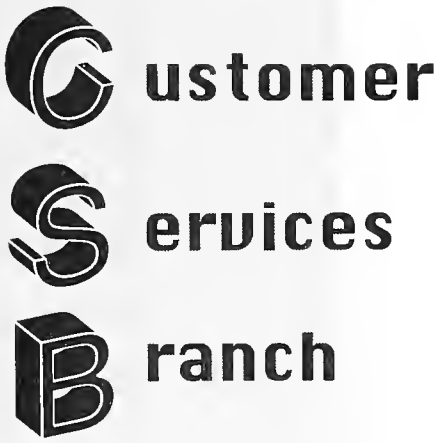
Future Plans

Massively parallel computing, workstation cluster computing, and other emerging technologies are continuing to be developed at the CFB. Within the next year, we will install a parallel supercomputer (SP/2) with several processors, to be used as an engine for selected, computationally intensive, scientific applications on the Convex.

Implementation of new communications front-end processors, which support high speed access to the IBM mainframes, should take place in the coming year.

Plans are being implemented to provide users with increased support for high performance scientific computing, improved security measures, and further development of ALW technology.

Technologies to support PCs and MacIntoshes® with technologies similar to those used in the ALW system will be investigated, as will technologies to provide central backup and archiving of LAN data.


Customer
Services
Branch

Customer Services Branch

Dale R. Spangenberg, Chief

The Customer Services Branch (CSB), which was established in FY93, provides centralized support services to the customers of DCRT, integrating the direct customer support functions previously scattered throughout DCRT. The Central Point of Contact Section, supervised by Dale Spangenberg, includes the Technical Assistance and Support Center (TASC), led by Geoffrey Marsh, and the Customer Accounts function, led by JoAnne Higgins. The newly established Training and Technical Information Section (TTIS), under Leslie Barden, encompasses the DCRT Computer Training Program and the Technical Information Office, which is led by Patricia Cleveland.

Highlights of 1994

This year has brought to fruition the plans and hopes of FY93, when the Customer Services Branch was established. Staff and contract resources were recruited from throughout DCRT. During the first half of the year, extensive preparations were made to set up service partnerships with the functions whose customer support would be assumed by the new Central Point of Contact. New contracts were established with Corporate Software, Inc., and Seneca, Inc., for extended PC and Macintosh® assistance. Tasks were defined, and contract personnel were brought in to conduct much of the direct customer contact. Development and training were done for Remedy, our help desk service tracking system. The available space was redesigned into an attractive and inviting area for walk-in assistance. A two-day team-building session helped bond the staff (who came from diverse computing platforms and diverse organizational cultures) and the contractors (from three different companies) into a mission-directed team. Finally, on January 31, the DCRT Technical Assistance and Support Center (TASC) opened.

TASC assumed support for the following DCRT services:

- PAL Unit Help Desk (IBM 370 mainframe assistance)
- PCB Help Line (PC , Macintosh®, and LAN help)
- DCRT Computer Training Program
- Project Control Office (Customer accounts)
- Statistical Support Staff

From the start, customers expressed their appreciation for these first steps toward "one call does it all," in particular for being able to get the answer to several short problems and register for a class in the same call. The choice of 4-DCRT (301-594-3278) as the telephone number for TASC has been very popular because it is so apt and easy to remember. Beginning in March, all calls to DCRT's SCRC (Scientific Computing Resource Center) were taken by TASC. On May 1, the Technical Information Office moved from the Computing Facilities Branch, and joined the newly formed Training and Technical Information Section in CSB. The next two months allowed time for cross-training and integration with the existing CSB services, while carrying out all continuing office functions. Finally, on July 1, the Technical Information telephone calls were directed to TASC, another major step toward the goal of providing a single point of contact for DCRT.

A regular program of feedback sessions has built camaraderie among the service partners throughout DCRT and allowed improved referral and tracking techniques to be shared by all. From the opening of TASC, customer satisfaction has been monitored through telephone surveys. Together with many spontaneous comments, notes, and electronic mail messages, the surveys attest to the success of the new unified customer service.

Three elements have been vital to the success of customer service within CSB:

- the multi-tiered call-handling structure
- participation in "Tier 1" by all staff
- quality of the participating staff

As developed by CSB, the tiered approach has all calls answered by a Tier 1 consultant, who is empowered to answer any question he or she can. Tier 1 consultants ask for assistance or hand on calls outside their areas of expertise to Tier 2 specialists within CSB, to subject specialists within DCRT, or to extended contract assistance. Tier 1 consultants have rapidly developed a wide acquaintance with DCRT resources and procedures for handling direct requests for assistance with accounts, computer training and technical publications, in addition to computer knowledge. The central place of Tier 1 in the customer services picture is insured by the direct participation by all senior members of the CSB as Tier 1 consultants for at least a few hours each week. All are kept directly in touch with the exhilaration and frustration of meeting customers' needs and expanding their own knowledge. Together with the rest of the DCRT team, TASC can provide meaningful assistance for more than 95% of customer requests. Of course, the quality of the

individuals taking on this challenge is critical to our success. In this we have been most fortunate.

The choice of the Remedy Action Request System as the Help Desk support tool has been vindicated, as this pioneering client-server tool has been implemented first in TASC and as a CSB management tool, and then in stages by the supported technical areas. Among the technical partners, Distributed Systems Branch took the lead, and from the opening of TASC has used the system to receive and return problem referral tickets and to generate a wide variety of reports for their management. Network Systems Branch shared their experience using Remedy to track trouble calls and let us use their server. The Statistical Support Section has used the system to submit tickets and accept and respond to referrals. More recently, the Computing Facilities Branch has begun to use a client station for management monitoring. Remedy's ability to serve UNIX®, Macintosh®, and PC clients allows the entire Division to participate in its use. Within CSB, Remedy is functioning on the desktops of all TASC consultants and managers on their personal Macintosh® and PC workstations.

Remedy allows consultants to review consults they have assigned to themselves or others, problems they have resolved in the past, and any requests or problems experienced by any particular customer. The value of these capabilities is increasing as the body of information in the "Customer" and "Service Ticket" databases grows and as the consultants and managers become more familiar with its use. Currently, the Service Ticket database contains information about more than 6,000 tickets. In July 1994, Remedy was installed on two new SUN® servers, providing both improved performance and the backup needed for this production system.

For the future, Remedy has the potential to allow customers to submit service tickets, monitor their status, and read responses as soon as these have been prepared. It can also serve as a knowledge base, allowing users to search for problems similar to those they are experiencing and to find their own answers. Many users will always prefer to call for assistance, but such a system can provide some assistance during "off" hours.

Central Point of Contact

The Central Point of Contact (CPOC) Section acts as the central clearing house within the DCRT for initial and follow-on user inquiries concerning DCRT services, acts as advocate for users to the DCRT and represents the DCRT to users. The CPOC also coordinates the various DCRT service user groups, and provides first level support to users on the correct use of DCRT facilities and clarification of problems and confusions. The CPOC provides consulting and resolution services for software and hardware problems encountered by users.

The following table shows the number of consults recorded in each platform category since the opening of the Help Desk on January 31 1994:

MVS	3105
PC	1442
Macintosh®	1063
Helix	391
UNIX® W S	23
DEC	9
Other	776
Total	6809

The CPOC also performs the following functions:

- refers complex user questions to, coordinates with, and participates in follow-on consultations with the appropriate Branches within the Office of Computing Resources and Services (OCRS)
- develops, maintains, coordinates and monitors an automated problem tracking system for use by all OCRS service organizations
- designs and develops methods for smooth handling of software and harware transitions within the OCRS
- generates reports regarding user contacts, user needs and user recommendations.

Technical Assistance and Support Center (TASC)

TASC receives most requests for assistance through calls to "4-DCRT" or e-mail to 4dcrt@nih.gov. During the year, the telephone lines formerly assigned to the diverse customer support functions at DCRT were redirected to the TASC number.

Systems Support

Development of the help desk support system with Remedy and enhancement and planning for the telephone system have been the major focus for FY94 . The robustness and performance of the Remedy system will benefit from the recent installation of two SUN® Microsystems SPARCstations as servers. In a series of meetings with NIH Telecommunications, Bell Atlantic and AT&T, the options available were studied, and the specifications were set for upgrades that will make it possible to count all incoming and outgoing calls, measure call durations, queue incoming calls, and perform many other functions. These

improvements will enable us to monitor our activities and serve customers efficiently.

Customer Accounts

During FY94, expanded responsibilities challenged the always busy customer accounts function. The growing popularity of ENTER SPONSOR gave account sponsors an electronic means of requesting changes to their accounts, so that the number of requests to be processed increased. Substantial time had to be devoted to planning for de-registration to improve security, especially for financial and administrative systems. Initiating cost recovery for the the new Helix and ALW systems added blocks of users to be registered, and re-registration for the Convex created an enormous and complicated re-registration process for all 2,000 members of that user community as the year drew to a close. Response time was temporarily increased, but soon returned to normal.

Fortunately, the future promises improvements. Additional staff and contract resources have been identified, and initial training of new employees to assist in these vital and complicated functions has begun. A new contract will be effective soon, and new support will become available by the end of FY94. Steps are being taken to reduce the duplicate efforts now needed to add new customers both on their individual systems and in the central database. In the longer term, a unified personnel and customer database system will enhance both the ease of updating and currency of the information for customer services.

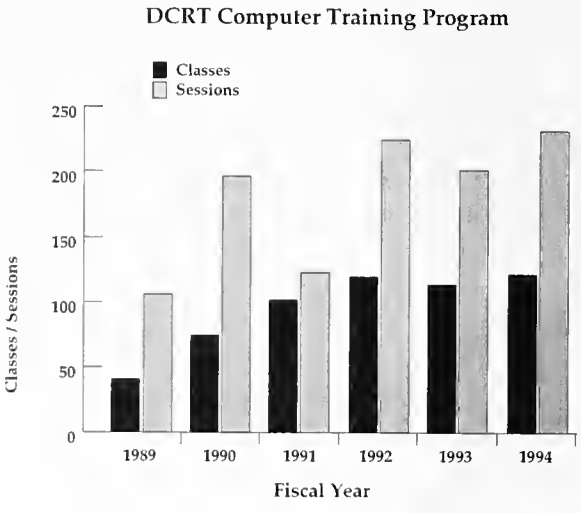
Training and Technical Information

The Training and Technical Information Section (TTIS) develops, manages and supports the DCRT Computer Training Program to enhance the ability of the NIH community to make effective and efficient use of computing resources and to make NIH aware of the expertise available at DCRT. TTIS keeps an extensive, current collection of computer documentation in stock at all times, and distributes publications and software upon request and by subscription.

Training

The many classes offered by instructors from throughout DCRT make the NIH community aware of the expertise available at DCRT. In addition, a variety of contract mechanisms are used to obtain instructors. Subject areas include personal computing, security, online mainframe services, database, networks and network services, statistical software, scientific computing, data analysis, molecular graphics, molecular modeling, and imaging. Various techniques are used to meet the needs of the diverse NIH computing community. Training approaches include planning and presenting formal classroom courses, brief seminars, and custom-tailored group presentations; enhancing and maintaining online training announcement and self-instruction systems; and evaluating, selecting, and distributing vendor-developed self-instruction programs.

During FY94, the DCRT Computer Training Program offered 232 sessions of 122 different classes. Growth of the training offerings in the last few years is shown in the following chart:



Beginning in January, students and instructors have enjoyed the results of an extensive renovation of the main classrooms in Building 12A. A new color video projection system was installed in the large lecture room, as was a new instructor demonstration station to allow real-time demonstrations on a 486 under DOS or Windows® and on a Mac Quadra®. The lab room was completely refurbished, with 12 pairs of new 486s and Macintoshes® at workstations sharing a single 17 inch color monitor. Each machine is provided with both network and modem connections, allowing instructors great freedom to choose the environment for class exercises. The renovated lab can accommodate 22 students in pairs at each workstation, with an additional workstation for the instructor.

This year the DCRT Computer Training catalog received an enlarged format and a new design. The familiar front-cover photograph was replaced with an image created by Peter FitzGerald on a Silicon Graphics workstation of a protein/DNA complex based on its crystal

Computer Training at NIH

Course Number Keys

4-digit	NIH Workforce Development Center	EPS	x56211
3-digit	DCRT Training Program	Building 12A	x43276
LIB-n	NIH Library	Building 10	x61156

Fall 1994

	DOS	Windows and OS/2	Macintosh	IBM MVS Mainframe	UNIX	General Concepts
Operating Systems	2502 Introduction to Personal Computing 2516 Introduction to DOS 6 0 2589 Intermediate DOS 6 0 835 DOS Batch Files 843 PC Troubleshooting 851 Memory Management on the PC	2590 Introduction to Windows 3.1 2598 Advanced Windows 801 Getting Started with Windows 821 OS/2 2.1 Overview 843 PC Troubleshooting	2785 Welcome to Macintosh 2764 Advanced Macintosh Techniques	723 Deregistration of Users for the NIH Computer Center	400 Fundamentals of Unix 408 Unix Commands 465 DCRT Support for Unix Workstations at NIH 700 Introduction to the Helix Systems	642 Supercomputing on the Internet 703 Central Computing Services at NIH 863 Introduction to Receiver Operating Characteristic Methodology 796 Perspectives on Management in the NIH Technical Environment: Organizing Chaos
Security	866 PC Viruses 2512 PC Disaster Recovery and Data Security	866 PC Viruses 2512 PC Disaster Recovery and Data Security		717 RACF for IBM MVS Data Security 718 Disaster Recovery	716 Security Lessons Learned in the Trenches	715 Introduction to Computer Security 721 Computer Data and the Privacy Act
Office Tools	826 PC Topic Sessions 878 Optical Character Recognition Technology 2585 Introduction to WordPerfect 6 0 2588 WordPerfect 6 0 - Advanced Topics 2517 Introduction to WordPerfect 5.1 2541 WordPerfect 5 1 - Advanced Topics 2515 WordPerfect 5 1 to 6 0 Transition 2577 Desktop Publishing with WordPerfect 5 1 2559 Introduction to Harvard Graphics, Rel 3 0 2565 Intermediate Harvard Graphics, Rel 3 0	826 PC Topic Sessions 2591 WordPerfect for Windows 2593 PageMaker 4 0 for Windows 2599 Harvard Graphics for Windows 4590 Freelance Graphics for Windows 4591 Desktop Publishing with WP for Windows 4580 MicroSoft Word 6 0 for Windows	2782 Introduction to WordPerfect 3 0 2784 Advanced WordPerfect 3 0 2780 Introduction to Microsoft Word 5 1 2781 Advanced Microsoft Word 5 1 2773 Introduction to PageMaker 2703 Quark Xpress 2774 PowerPoint 2796 Aldus Persuasion 2792 MacDraw PRO 1 5	152 Introduction to WYLBUR 156 Beyond Basic WYLBUR 2630 IMPACT System for Personnel Staff 2631 IMPACT System for Management (MSCs) 2632 IMPACT System for Administrative Staff 2633 IMPACT System for Professional Staff 2634 IMPACT: A-Train 2580 Introduction to CRISP 2581 Advanced CRISP 2582 CRISP Thesaurus		775 An Introduction to Neural Networks LIB4 Network Access to Physicians' Desk Reference
Database	2506 Introduction to dBASE III Plus 2508 Intermediate dBASE III Plus 2583 Introduction to dBASE IV 2584 Advanced dBASE IV 2539 Introduction to Symphony 2540 Intermediate Symphony 2575 Advanced Symphony 2552 Introduction to Paradox 2561 Advanced Paradox LIB6 Grateful Med Setup, Loansome DOC LIB7 Using MeSH with Grateful Med	4597 Access for Windows 4594 Introduction to Paradox 4 5 for Windows	2790 Introduction to FileMaker Pro 2704 Advanced FileMaker Pro 2770 FoxPro- Levels 1 & 2 LIB6 Grateful Med Setup, Loansome DOC LIB7 Using MeSH with Grateful Med	312 OMF DB2's Query Management Facility 315 DB2 Database Administration 316 DB2 Application Programming 318 QMF Access to Human Resource Data for Personnel Staff 319 QMF Access to Human Resource Data for EEO Staff 309 ADBIS AdHoc Query Using OMF 322 Client - Server Access to Centrally Managed Data 325 Database Technology Series 329 Graphical Access to the ADB and ADBIS	466 Andrew File System for Advanced Labdratory Workstations	301 Relational Database Overview 308 SQL: The Language for Relational Databases 313 Relational Database Design
Spreadsheets	2504 Introduction to Lotus 1-2-3 Release 2 4 2505 Lotus 1-2-3 Release 2 4 - Advanced Topics	2592 Lotus for Windows 2595 Excel for Windows 4501 Quatro Pro for Windows	2787 Introduction to Excel 4 0 2789 Advanced Excel 4 0			
Statistics		222 Preview of the SAS System for Windows 226 Easy Statistical Software for Windows Overview 227 Easy Statistical Software for Windows Hands-On Lab	2776 KaleidaGraph 3 0	212 SAS Fundamentals I for Non-Programmers 212 SAS Fundamentals II for Non-Programmers 215 SAS Fundamentals I for Programmers 216 SAS Fundamentals II for Programmers		203 Statistical, Mathematical, and Graphics Software Fair 250 Basic Statistics Using SAS/STAT 251 Using SAS/STAT Procedures to Perform ANOVA and Regression 957 MATLAB 990 Tools for Statistical Inference: Methods for Exploring Likelihood Functions and Posterior Distributions
Mail, Networks and Connections	321 Minuet for PC DOS Network Access 622 Using GOPHER on the PC 857 Electronic Forms on PUBnet 2801 MS Mail for DOS	622 Using GOPHER on the PC 2805 MS Mail for Windows	359 Macintosh Networking with TCP/IP 621 Using GOPHER on the Mac 2803 Eudora Mail for Macintosh 2807 MS Mail for Macintosh	760 ENTER MAIL 770 ENTER BBS 362 LISTSERV Electronic Mailing Lists 750 Kermit 752 ProComm Plus	364 Using the Internet 630 Automated Information Retrieval Using Autogopher	350 Introduction to Networks 355 LAN Concepts 356 Using Netware at NIH 357 Planning and Installing a LAN 612 Network Services 368 Advanced Network Topics 624 The World Wide Web, Mosaic, and NIH
Programming	865 Introduction to the BASIC Language 2562 Paradox PAL (programming)		LIB3 Searching MedLine Using Command Language	270 COBOL/370 Conversion		789 System Modeling for Application Development
Scientific Applications	980 Analysis of Ligand Binding Data Using the Ligand Program 863 Introduction to Receiver Operating Characteristic Methodology LIB1 Searching MedLine with Grateful Med		860 Drawing Pedigrees with PedDraw for the Macintosh, Hands-on Laboratory 962 Image Processing on the Macintosh 965 Using Image 1 56 for Densitometric Analysis of 1-D Gels 980 Analysis of Ligand Binding Data Using the Ligand Program LIB1 Searching MedLine with Grateful Med LIB2 Managing References with Macintosh Software		570 GCG Sequence Analysis 574 Genomics Software for Unix Platforms	640 Using the Internet for Sequence Analysis 900 DCRT Speaker Series 903 Scientific Data Analysis: Methods 905 Scientific Data Analysis: Resources at NIH 927 Electrical Noise as a Tool for Molecular Kinetic Studies LIB5 Molecular Sequence Information Using Entrez

structure. This change highlighted the increasing number of scientific seminars in the training program.

In addition to classes for the DCRT customer community, CSB organized a number of classes intended primarily for DCRT staff. During FY94 these included ObjectView from KnowledgeWare, Inc., Software Systems Analysis and Design from Learning Group, Inc., GUI Design from PowerSoft, Inc., Visual BASIC for Developers from Advanced Paradigms, Inc., and DFS System Administration from Transarc, Inc. Bringing programs such as these to our own classrooms facilitates and expands employee training, while reducing travel costs.

Technical Information Office

The Technical Information Office is responsible for the acquisition, inventory, and distribution of vendor documentation, various DCRT publications and certain software to the DCRT customer community. The Unit maintains an inventory of 1,450 different publications, with an average of 90,000 publications on hand. It provides a basic distribution service for all DCRT mailings by administering databases and generating mailing lists. An individual documentation profile is maintained in an online database for each customer who receives documentation. Profiles are updated whenever new publications are requested and when the customer asks for them to be discontinued. Subscription lists are renewed annually, and those not renewing are dropped. Documents can be ordered through an online ordering system from the user's terminal or personal computer or by telephone or walk-in request. During FY94, 5215 individual orders were processed, and a total of 118,994 publications were sent to customers through the subscription service or in response to requests.

Future Plans

During the next year, CSB looks to strengthen existing services and continue expansion into additional service areas, as we progress towards our goal of being the central point of contact for all DCRT support services. Cross-training of CSB staff will allow for greater support of each service area and flexibility to shift resources in response to urgent needs. We also look forward to additional staff resources from new contracts. The tasks assigned to the new staff will provide additional support for TASC, customer accounts and systems development. Technology improvements will also play a role in helping us achieve greater efficiency and effectiveness. Phone system improvements will allow us to respond more efficiently to peak call volumes and capture statistics to assist in resource planning. Redesign of some of the systems and processes associated with customer accounts will streamline our registration processes and unify our customer data.

As CSB grows, we look forward to playing a key role in advocating DCRT customer needs to DCRT management. In our Remedy database, we are building a wealth of information, which will help us identify trends in computing at NIH and help us adjust services to meet changing customer needs. Efforts are under way to standardize the information content in Remedy and to build meaningful analyses, which will improve our ability to identify DCRT service needs.

Distributed Systems Branch

Distributed Systems Branch

David Songco, Chief

The Distributed Systems Branch was formed in 1993 as part of the DCRT reorganization. Most of the staff and functions of the Personal Computing Branch were combined with components of the Computer Systems Laboratory, as well as with staff from two other labs, to form the DSB. This integration brings together nearly two dozen highly skilled computer specialists, whose mission is the ongoing support of PC, Macintosh®, and LAN users, with about a dozen senior engineers, programmers, mathematicians, systems analysts and scientists, whose mission is the design and development of laboratory and clinical systems in support of biomedical research. In addition, to bring particular emphasis to the special computing needs of NIH scientists, the management of the DCRT Scientific Computing Research Center was formally assigned to the DSB.

DSB serves as the DCRT focal point for the development, support, guidance and application of local area network, workgroup, and other forms of distributed computing at NIH. The DSB mission includes advocacy of, advice about, and assistance with the spectrum of computing platforms normally found in the office, laboratory, or clinic. The DSB is responsible for assessing the technical requirements of the user community in the area of distributed computing technology and for ensuring that those requirements are addressed in the future goals of the DCRT. DSB:

- Promotes the effective use of distributed systems technologies at the NIH by fostering working partnerships with NIH ICDs
- Provides technical guidance, education, and publications on relevant hardware, software, and overall system architecture
- In collaboration with ICD scientists, develops computer based systems that have potential for wide application in biomedical research at NIH
- Provides the NIH community with resources for high-level computational analysis of data in the field of molecular biology
- Provides guidance in the selection and effective use of computer image technology in the clinical and medical informatics areas
- In conjunction with other DCRT labs and branches, collects, develops, and distributes software utilities of general use to the NIH community
- Operates the Scientific Computing Resource Center and sponsors other NIH User Resource Centers (URCs) as part of the consulting and education function.

DSB Organization and Staffing

The DSB comprises the *Office of the Chief*, four sections, and the Scientific Computing Resource Center. Current staffing levels include 30 government positions, 6 part-time student employees, and 11 contract staff. The DSB use of contract staff has increased this year to offset the loss of staff during a period of hiring restrictions.

The Office of the Chief is responsible for overall branch planning, management, and policy. It also provides selected mathematical and statistical consulting, as well as project management for DSB and DCRT collaborative projects.

The *System Consulting Section* (SCS) assesses the technical requirements of the NIH user community in the area of distributed computing, including PCs, Macs, workstations, LAN applications, and associated technology, and provides guidance and support for NIH organizations in the selection and effective use of emerging workgroup computing technology in the laboratory, clinic and office environment. It sponsors and coordinates the NIH User Resource Centers in collaboration with the NIH Training Center, and develops and delivers training in collaboration with the DCRT and NIH training programs. SCS staff manage the NIH Computer Support Coordinator program and coordinate the dissemination of information to the NIH user community.

The *LAN Technology Section* (LTS) tracks and evaluates emerging local area network computing technologies, including network operating systems (NOS) and enterprise e-mail systems. LTS staff develop LAN system models and recommendations for technical solutions to common NIH LAN computing environments, administer the Network Operating System (NOS) and electronic mail components of the DCRT local area networks, and collaborate with the Network Systems Branch in developing campus-wide network services.

The *BioInformatics and Molecular Analysis Section* (BIMAS), headed by John Powell, provides the NIH community with resources for high-level computational analysis of data in the field of molecular biology and genomics. This section also provides guidance and support in laboratory data management and in the use of various software/hardware packages, and it establishes ties to other DCRT and NIH offices, as well as to national and international resources.

The *Laboratory and Clinical Applications Section,* headed by Kenneth Kempner, performs requirements analyses for laboratory and clinical application areas. It tests, evaluates, designs, develops and implements computer systems and software applications in support of laboratory and clinical research programs.

The *Scientific Computing Resource Center (SCRC),* under the direction of Brian McLaughlin, provides NIH with a shared-use computing facility, staffed by computer professionals, where researchers are able to focus on scientific applications. It addresses many of the needs of the NIH scientific community by providing access to scientific software running on advanced personal computers and UNIX® workstations, and fosters scientific computing by offering scientific software in specific subjects such as image processing, molecular modeling, numerical analysis, sequence analysis, and statistics.

Distributed Computing Technology Highlights

Architectural Management

A major aspect of our technology tracking efforts this year was our participation in the DCRT-wide cooperative effort to establish a technology architecture for NIH. The goal of this project is to achieve consensus on standards for a technical architecture that will meet NIH and ICD scientific and business requirements for computing and networking. Choosing a consistent technical architecture can help NIH to substantially reduce training and support costs, reduce the burden of tracking technology and upgrades, improve communications among the various platforms at NIH, and cut wasted time and effort for file and format interchange.

DSB staff provided the technical leadership for establishing standards for the PC and Mac clients as part of an overall client/server architecture. DSB also focused on integrating an e-mail and NOS strategy into the overall architecture.

Desktop Hardware

For PC users, the DSB currently recommends Dell microcomputer systems for most applications, but continues to look at alternatives, especially the offerings of Compaq. As for the IBM line, we are no longer recommending the Microchannel architecture.

During 1994, prices dropped dramatically on 486 and Pentium systems. Thus, we began recommending Pentium systems for power users, 486/33 machines for entry level and 486/66 computers for mainstream use. Price now makes Pentiums a good value for users with moderate to heavy business needs, although they do not bear Energy Star certification. We hope to see the first Energy Star Pentiums in the winter of 1994/95.

For Macintosh® users, FY94 marked the introduction of another innovative new computer platform called "Power Macintosh®" to mark the 10th anniversary of Apple's Macintosh® computer. These machines are built around the new PowerPC microprocessor. This chip, developed collaboratively by Apple, IBM and Motorola, utilizes RISC technology to achieve high performance in a small, low-power chip.

Because the PowerPC processor uses a different instruction set from the Motorola 680x0 chips upon which older Macintosh® models are based, software needs to be rewritten to take advantage of the speed of the PowerPC. In an extraordinary technical achievement, Apple produced a software emulator that allows older Macintosh® software to run on the new PowerMacs, albeit at reduced speed.

In addition to running Macintosh® software, the new PowerMacs can also run most DOS and Windows software by using an emulator called "SoftWindows." Produced by Insignia Corporation, it runs DOS and Windows programs at 486 speeds. Unfortunately, at this time, SoftWindows only emulates the instruction set of the Intel 80286 chip.

The DSB conducted a study of energy–efficient computers during 1994, providing guidance and recommendations for implementing the Executive Order to purchase Energy Star microcomputers.

Desktop Software

The DSB increased its efforts to move DOS users to the Windows environment, and this effort appears to be reasonably effective; as 286 systems are replaced with 486 systems, Windows becomes practical. Windows for Workgroups is becoming the dominant seller, due to the proliferation of bundling and networked environments. DSB staff have remained active beta testers of new Windows versions, including NT 3.5 and the next mainstream Windows version, code named "Chicago." Sales of OS/2 continue, but its future appears to be at risk due to its difficulties in providing support for Windows for Workgroups and "Chicago."

Our specialists remain very interested in product suites, especially Microsoft Office, and groupware applications, such as group schedulers and Workgroup

Information Manager, CaLANder and ECCO 2.0. This year, DSB staff implemented electronic scheduling of conference rooms for DCRT.

Since the introduction of the Power Macintosh®, software developers have been scrambling to rewrite their applications to take advantage of the speed of the PowerPC chip. This requires re-compiling all 680x0-specific code to work with the PowerPC. Such applications are referred to as "PowerPC Native."

In some cases, developers have released "Fat Binary" versions of their programs, which can run on either 680x0 or PowerPC platforms in native mode. The DSB anticipates that most vendors who make the switch to PowerPC native applications will also continue to support older models of Mac. A few select applications that require a high-performance processor, such as mathematical and molecular modeling programs, will probably be available in PowerPC-only versions.

Support and Guidance for Locally Managed LANs

DCRT Local Area Network (LAN)

DSB staff successfully migrated the DCRT LAN from the 3Com 3+Share Network Operating System (NOS) to Microsoft LAN Manager and finally to Microsoft's latest networking product, Windows NT Advanced Server (NTAS). In addition, DCRT changed electronic mail products. The LAN changed from 3Com's protocol-dependent e-mail (3+Mail) to Microsoft's protocol-independent offering, MS Mail. A number of Macintosh® users have elected to implement Eudora POP mail software, taking advantage of the NIH-wide POP server. DSB has provided training sessions for DCRT users on the use of these new products. We plan to assist other groups on the NIH campus now undertaking a similar migration to the newer technology.

Microsoft Windows NT Advanced Server (NTAS)

In collaboration with the Network Systems Branch, DSB provided numerous organizational consultations to assist ICDs in their LAN planning. DSB has been assisting the Technical LAN Coordinators (TLC's) in resolving problems arising during their migration from Microsoft LAN Manager to NTAS. DSB also renewed the "Microsoft Premier Support Contract," which provides unlimited direct support from Microsoft for NTAS and many other Microsoft products. This contract also allows the NIH ICDs to take advantage of this service at a greatly reduced rate. DSB continues its leadership position in tracking the next generation of NTAS (Windows NT Server 3.5).

Novell NetWare

This year, in response to growing interest at NIH, DCRT began limited support for Novell NetWare 4.x. To this end, NSB is now routing the NetWare protocol, IPX, and DSB staff have received training in NetWare 4.x and have hired contractors with NetWare experience. Together with the contractors, DSB has been installing and evaluating a test NetWare environment. Particular attention is being paid to how well NetWare integrates into the NIH environment. DSB and NSB are meeting with NIH NetWare administrators to discuss implementation issues for an NIH-wide Novell Network. DSB's support is focused on NetWare 4.x. and the migration from version 3.x to 4.x. Site licenses and support mechanisms are being investigated.

DECnet

The number of DECnet nodes on the expanding NIH campus-wide network continues to climb. An increasing number of VAX sites use DECnet networking protocols to support personal computing via DECnet-based server configurations. Nearly three quarters of the DECnet nodes now registered are personal computers, using work group servers for file and print services. During the coming year, we will continue to provide the centralized coordination necessary for smooth incorporation of new DECnet hosts into an integrated network, which spans the entire NIHnet campus network.

Electronic Mail

Electronic mail has become increasingly important at NIH. Joe Murphy and John White have been leading our efforts to provide comprehensive e-mail capabilities for the many PC and Mac users at NIH. During FY94, DCRT continued to promote and support MS Mail as the successor to the 3Com mail at NIH. The DSB is the central distribution point at NIH for those LANs that wish to install MS Mail. In addition, DSB provided guidance to LAN administrators converting from MS Mail or installing electronic mail for the first time on new LANs. There are now close to 80 LANs running MS Mail, nearly three times as many as at this time last year. The DSB also continues to work with the Network Systems Branch in developing and maintaining the central NIH Microsoft Mail hub and Simple Mail Transfer Protocol (SMTP) gateway.

The DSB, in collaboration with CFB, supports Postoffice Protocol 3 (POP3) mail for users at NIH

who wish DCRT to provide central e-mail service so that they need not maintain their own mail servers. DSB works with the CFB to promote POP3 usage and provide end user support for the POP3 e-mail client, Eudora. DCRT plans to expand our POP3 capability substantially in FY95, so that many more users can access central e-mail services.

Much attention was paid to Transmission Protocol/ Internet Protocol (TCP/IP) networking issues in FY94, particularly with regard to TCP/IP on Macs and PCs. New versions of TCP/IP drivers and applications programs were tested, and DSB provided leadership and consultation to end users on the use of TCP/IP and access to TCP/IP network services such as GOPHER and MOSAIC.

Workstation support also focused on the new versions of System 7 and Windows for Workgroups, which allow workstations to share data through the network more easily. Macintosh® users in particular have benefitted from new versions of system software that DSB has tested and made available on PUBnet.

New LAN Technology Challenges

As in all areas of computing, there are many challenges ahead in the networking arena. DSB wants NIH to be in the strongest position to take advantage of changes in technology. The role of standards in computing must not be overlooked.

We are evaluating the tradeoffs between a standards based e-mail strategy versus a vendor driven approach, such as Microsoft's EMS, and whether NIH should limit itself to one Network Operating System campus-wide. There are a number of issues that continue to surface with the implementation of more than one NOS at the NIH. Support, service, training and interoperability all must be addressed, as well as the cost/benefit of having multiple e-mail and NOSs to the individual ICDs as well as to NIH as a whole.

Other questions being addressed concern the role that the Distributed Computing Environment (DCE) should play in the future of networking, how DCE affects networking at NIH, what steps NIH should take to insure that we will be prepared to take full advantage of DCE, should it become "the" standard, and the future of Microsoft and Novell in this arena.

Training

Quality support begins with quality training. The DSB training program, coordinated by Rick Duhn, provides training development and delivery via its collaborative relationships with the NIH Training Center and the DCRT training program. In the DCRT program, there has continued to be interest in the advanced DOS courses (DOS Advanced Topics; Batch Files), and an increase in two or three hour limited–focus seminars. Seminars of particular interest to the scientific community include:

- Image Processing on the Macintosh®
- DNADraw for the Macintosh®
- Analysis of Ligand Binding Data
- Pedigree Drawing Programs for the Mac and PC
- Running SAS Software in the PC Environment
- Preview of the SAS System of Windows
- SAS/JMP for the Macintosh®
- BASIC Biomedical Computation

Seminars of general interest to computer users include:

- Mac and PC Viruses
- A Look at DOS 6.0
- Memory Management on the PC
- PC/Macintosh® Data Exchange
- Macintosh® Networking with TCP/IP
- Windows Applications Strategies
- PC Troubleshooting
- OS/2 Overview

There has been a particularly strong interest in the "Getting Started with Windows" seminar, as more ICDs upgrade their hardware to handle the Windows environment; DSB arranged and taught a number of these seminars on location for specific ICDs such as DRG, NICHD, and NCI.

At the NIH Training Center, there has been a continuing, but somewhat diminishing, interest in DOS–based courses (WordPerfect; Lotus 1–2–3; dBase III and IV; Paradox; Disaster Recovery), while Windows–based courses have had an increasingly strong showing, (e.g., WordPerfect; Lotus 1–2–3; Excel; Harvard Graphics; PageMaker; Microsoft Project).

Although interest in Macintosh® applications courses continues (WordPerfect; Microsoft Word; Excel; Lotus 1–2–3; FileMaker; PageMaker; Quark Xpress; MacDraw; KaleidaGraph; PowerPoint; HyperCard), the number of participants has declined compared with previous years. Interest in end-user networking and e-mail courses (MS Mail and Eudora) is moderately high.

In FY94, 219 personal computer course sessions were presented by the NIH Training Center and attended by 2,573 students. Additionally, 66 course/seminar sessions comprising 1,596 students were presented through the DCRT training program without fee. In all, DSB or other DCRT staff taught 23 percent of all courses and seminars offered, while outside vendors,

many under the direction of the DSB and using DSB-developed materials, taught the rest.

Interestingly, while the total number of courses was lower than the previous year (285 in FY94 vs. 310 in FY93), more students were taught (4169 in FY94 vs. 4018 in FY93).

Organizational Consulting

During 1994, the DSB substantially shifted the focus of its consulting resources to assisting organizational groups rather than individuals. We provided guidance and assistance to ICD staff who make strategic distributed computing decisions for their organizations and to those who provide first-line support. Assistance is provided in selecting hardware systems and application software at the organizational level, implementing local area networks, devising cohesive database management strategies, and developing local support capabilities. We put substantial effort into developing and promoting interoperability among the computing platforms at NIH.

As part of this strategy, DSB encouraged enhanced local support by sponsoring the newly formed Computer Support Coordinator program, an amalgam of our old Lead Users and Macintosh® Support Coordinators.

In moving toward organizational consulting, more of the responsibility for direct end-user problem resolution is being provided by the newly established DCRT Customer Service Branch, which in turn leverages its support by outsourcing many of the routine requests for assistance. In this scheme, DSB staff serve as third tier consultants for the more complex and NIH-specific problems.

An example of our organizational consulting is our collaboration with the NIH Office of the Director (OD) LAN Support Center. We are working closely with this group to develop and implement a plan to migrate over 1,200 users from 3Com based LANs to Microsoft NTAS. We are also helping them develop a new e-mail strategy and a plan to move their DOS users to Windows or Macintosh® technology.

Another example of how we help groups use computers more effectively is our work with one of the NCI research labs, where we were asked to review and evaluate the lab's administrative processes, with particular emphasis on the electronic work flow process of scientific manuscript production. DSB staff reviewed such issues as electronic bibliographical reference management, staff access to Medline, hardware requirements for adequate data storage, standardization on procedures and software used in manuscript production, and e-mail and Internet access for scientific and support staff. Specifications were developed for new hardware and software, and initial training was provided to NCI staff members.

Network Consults

Network consults continues to be our most active organizational consulting activity. We provide guidance and recommendations for desktop operating systems, network operating systems, LAN software such as server backup software and e-mail, and other workgroup technologies. We collaborate with other DCRT components for network cabling and access to NIH services.

DSB Partnership with the Users

Computer Support Coordinators (CSC)

DSB has formed a partnership with ICD local support staff in order to serve over 16,000 PC and Mac users at NIH. To increase the efficiency and effectiveness of our support, DSB combined the Lead User and the Macintosh® Support Coordinator programs into the CSC program. The CSC program now serves about 125 individuals, who attend monthly meetings to discuss support-related topics, participate on LISTSERV for timely dissemination of information, and receive special training opportunities. CSCs enjoy training benefits aimed at enhancing local support. DSB staff presented a number of free CSC seminars on such topics as hardware troubleshooting, memory management, and file recovery. The DSB also sponsored vendor training on Microsoft NTAS for CSCs.

NIH PC Topic Sessions

Consequent to the formation of the CSC Program, we separated out the monthly Lead User Topic Sessions as NIH PC Topic Sessions open to the NIH public. Presenters so far have been Dell and Compaq, Microsoft on Microsoft Office, and Adobe on their multimedia product, Premier.

Associate Instructor Program

The DSB Associate Instructor Program continued to play a significant role in the success of our training efforts, and a special recognition ceremony was held in June, at which Associate Instructors were honored for their many years of faithful service. Under the program, experienced NIH computer users volunteer their time

assisting the primary instructors during hands-on training courses. During FY94, 120 persons from across NIH's ICDs and the Office of the Director participated in this very successful program.

Biomedical Research Macintosh® Users Group

The Biomedical Research Macintosh® Users Group (BRMUG) featured monthly presentations on scientific and administrative topics, from gene sequencing programs to exchanging data between Macs and PCs. BRMUG once again sponsored the BRMUG Scientific Show, which included 20 scientific hardware and software vendors.

Wordprocessing Users Group (WUG)

NIH employees meet once a month to present a variety of word processing related products and to discuss issues related to writing with a computer. Individual group members volunteer to demonstrate their favorite features at each of the meetings. At least once each year, two group members contrast and compare two different packages (e.g., WordPerfect vs. MS Word). Occasionally, major vendors are invited to reveal the changes in their new software releases.

Campus Users Resource Exchange (CURE)

The CURE meets monthly for network related presentations and information exchange. CURE also serves as the forum for the NIH Technical LAN Coordinators. FY94 produced many lively discussions on the future directions for network operating systems and e-mail at NIH.

Partnership with Industry

DSB staff participated in the MacWorld Advisory council and presented workshops at MacWorld in Washington. DSB staff also participated in both the Fall 93 and Spring 94 Interop meetings as well as Fall and Spring COMDEX.

Information Dissemination

A central element in the DSB's mission continues to be the dissemination of information about distributed computing. The DSB information dissemination program is coordinated by Daniel Zoll. During the past year DSB produced five issues of PCBriefs, each issue being distributed to about 7,000 NIH employees. As our "flagship" publication, PCBriefs contains such elements as technology and product reviews, training schedules, software and hardware upgrade information, tips on hardware and software and "Q&As."

DSB publishes PC and Mac "Product Information Guides" (PIGs). While each has been updated at least twice during the past year, product lines are changing so fast that it is a major task to re-edit and re-print these increasingly large publications. Thus, we have made a decision to break the PIGs into discrete parts and publish these parts separately as DSB "White Papers." These papers will be available electronically over PUBnet and through the User Resource Centers, and in hard copy through a request to the Technical Assistance and Support Center (TASC) unit of CSB.

This year the DSB published A Cross-Platform Computing Guide, a detailed guide for dealing with document interchange in cross-platform environments, written by Mike Basham; Kevin Haney produced a paper on Windows NT Advanced Server Security. Dale Graham of the SCRC produced new handouts on Pedigree Drawing and Creating Internet Documents and revised a number of other documents including: Comparing Sequence Analysis programs; Getting and Finding Sequences on Molecular Biology Databases; and Multiple Sequence Alignment on the GCG. She also presented a poster on poster making at the NIH Research Festival.

DSB has also continued to support PCBull, its electronic bulletin board, which makes available conferences (interactive questions and answers), latest training schedules, downloads of helpful files, the latest PC information and DSB publications. PCBull, however, is strictly a modem-access BBS, and this is increasingly a limitation in the NIH environment. Thus, we are currently beta testing a new LAN based BBS on Macintosh®, code named DSBLink, and built on First Class software. If this beta test proves successful, DSBLink will open up to both Macs and PCs and also support dial-in access. We envision that it will replace PCBull.

Under the leadership of Catherine Greenville, the DSB provides electronic government forms in various application formats as a service via PUBnet. In order to make these forms available to the largest group possible, DSB has made them available to users of AppleTalk, 3COM, and LanManager networks. Currently available are forms created with File Maker Pro, Form Flow, and WordPerfect. These forms can be used by Macintosh® or PC computer users who have a copy of the application software. DSB does not create forms, but if users have created forms they wish to share, DSB will make recommendations and clear the forms for use by the NIH Forms Manager.

PUBnet

The NIH Public Network is a network facility, developed and managed by Lori Collins and Jim Del Priore, that provides computer related information, software and tools to users of NIH LANs. DOS, Windows, and Macintosh® users can easily connect to these services from their native desktop environments in order to download software and information or run tasks directly. In 1994, the PUBnet project team and the NIH Library collaborated to provide a smart e-mail system called PUBnet Auto Mail (PAM). This system allows users to request PUBnet documents and files through e-mail requests. In the coming year the PUBnet project team will strive to provide native access to Novell Netware workstations as well as some popular Internet interfaces such as MOSAIC.

Security

The DSB Security program, led by Kevin Haney, provides guidance and support in the areas of microcomputer security, viruses, safe computing practices, and data recovery. DSB offered a class on PC viruses, and continued support for our site-licensed antiviral software for PC and Macintosh®. While small numbers of viruses continue to be found on campus, no extensive viral outbreaks were reported during the last year.

Computer security took on more of a Division-wide focus this past year. The DSB Security Coordinator and several other DSB staff were instrumental in the formation of the DCRT Computer Emergency Response Team (CERT), whose purpose is to provide coordination and communication among DCRT staff during incidents of system or network intrusion of DCRT computing systems. The DSB Security Coordinator is currently serving as CERT Coordinator. The DCRT CERT coordinated the response to the Internet "sniffing" incidents and hacker attacks during May and June, and worked with the NIH Office of Information Resources Management (OIRM) and various law enforcement agencies in the investigation of these and other incidents. The CERT Coordinator also organized two public discussion panels on the sniffer incidents.

DSB staff participated in a security review of DCRT by the Carnegie Mellon CERT (the Federally-funded group responsible for overall Internet security). This review will provide DCRT with recommendations on improving the overall security environment of DCRT systems and should also help the division to offer more secure systems, resources, and services to campus users.

Other security activities engaged in by DSB staff include collaborating with the NIH OIRM to conduct computer security awareness programs for DCRT and NIH employees, and working with OIRM to develop a network risk assessment methodology for NIH LANs. This novel methodology will enable NIH LAN administrators to assess the security levels of their PC-based LANs and provide recommendations for improvement where necessary. The first DCRT class in network security was offered during the summer, and was taught by DSB staff. The DSB Security Coordinator participated in several NIH investigations that involved breaches of computer security and data recovery.

User Resource Centers

In collaboration with the NIH Training Center, the DSB is co–sponsor of two User Resource Centers (URC), one on the main campus, and one on the Executive Boulevard corridor. Both are multipurpose walk–in computer facilities, equipped with PC and Macintosh® workstations, an extensive collection of applications software, and a variety of peripherals, such as laser printers, modems, CD-ROM players and page scanners. They also have a large selection of self-study courses and many popular personal computing periodicals, books, catalogs, and other publications. The NIH URCs have continued to serve as a vital adjunct to DSB training services.

During FY94, NIH employees made more than 7,100 visits and 1,200 telephone calls to the URCs for such purposes as researching computer topics, evaluating DSB-supported hardware and software, taking self-study courses, and consulting with URC staff.

The URC Learning Assignment Program, under which NIH employees volunteer 4 hours of their time per week for 3 months in exchange for the opportunity to enhance their computer skills by working directly with URC staff, remained popular.

As part of DSB's collaboration in the URC, DSB staff rewrote, using FoxPro, the URC's seriously outdated and outmoded login program for both Macintosh® and PC.

Animal Information Management System

Under the leadership of Robert Romanoff and Ramon Tate, DCRT and NCRR/VRP began a collaborative Animal Information Management System (AIMS) project in FY94. The overall objectives of the project are: to determine what is needed to provide NIH with tools for animal data management; to determine

data needs for research, service programs, management decisions and compliance with regulatory requirements related to animal research conducted by the NIH intramural programs; to develop and promote standards for data storage and sharing; and to produce a requirements analysis and design for the next-generation animal information system in the Veterinary Resources Program (VRP), NCRR.

VRP is the largest centralized research animal care and use program in DHHS. VRP's areas of animal data management involve elements that are needed by NIH as a whole, but with a different emphasis for each ICD. Thus, the needs for all of NIH will be solicited and included in the requirements for the VRP system. All levels of the NIH community of researchers and animal facility personnel are being asked to help with this endeavor. The resulting system will be a model that will be attractive for other ICDs to duplicate and/or use. A major goal will concentrate on payback to researchers by providing more electronically accessible information about animals used in their protocols.

In FY94 a project plan was formulated by the AIMS team and agreed to by the Directors of NCRR and DCRT. Appropriate standards, system engineering tools, and methodologies were selected, and a short-term training project using these tools and methods has been completed. Development of a business model of the VRP system requirements was started, as was evaluation of the feasibility and resource requirements of various short-term projects.

By the end of FY95 or early FY96 the strategy and analysis phases of AIMS should be complete, resulting in comprehensive requirements specification for the VRP system. In FY95 we also expect to complete some of our short-term AIMS projects.

Scientific Computing Technology Highlights

Scientific Computing Resource Center

B. McLaughlin, Ph.D.
with D. Chow, D. Graham, Ph.D., F. Marsh, S. Schien, R. Tate, Ph.D., J. Tomlin

The Scientific Computing Resource Center (SCRC), which opened in May 1992, is a shared-use computing facility, where NIH researchers are provided hands-on access to powerful desktop computers running a variety of scientific software packages. SCRC users are encouraged to evaluate the utility of a computational program or tool by using it to analyze real data from their own research. SCRC staff and other DCRT consultants are available to provide guidance in the effective use of SCRC advanced personal computers and UNIX® workstations and computational tools, with emphasis on software for image processing, molecular modeling, sequence analysis and statistical data analysis.

The array of SCRC hardware user workstations consists of Macintosh® Quadras, 486 PCs running Windows, Silicon Graphics Indigos running UNIX®, and one SUN Sparcstation configured as an Advanced Laboratory Workstation (ALW). SCRC workstation users have access to over 75 SCRC scientific analysis software packages. In addition, SCRC facilities include high-resolution image acquisition tools (scanners, digital cameras, film scanners) and high-quality output devices for use with the SCRC scientific software.

The goal of the SCRC is to promote the evaluation and use of software in the analysis of scientific data. Any NIH employee may use the scientific resources at the Center for evaluation purposes or for an occasional short–term project. The SCRC tries to put different types of scientific computing solutions into operation, so that researchers can make informed decisions about which resources are most needed in their lab or office.

SCRC Passes the Two-Year Mark

The assessment of the SCRC's year-long pilot phase resulted in the SCRC re-sharpening its focus on software for image processing, molecular biology, molecular modeling, and statistical data analysis. For the first time, all users were asked to call ahead for SCRC appointments so that appropriate resources could be made available systematically. It became necessary to place use limitations on SCRC scanners and the highly popular SCRC high-end color printers. The SCRC regretfully decided that this is not a hands-on technical graphics production facility. Although such a service would be widely useful, it is beyond the scope of the current SCRC resources. The charter of the SCRC is to provide the capability and support for researchers to use and explore scientific software tools in analyzing ongoing NIH research problems.

During the past year, James Tomlin and the other SCRC staff were visited by about 550 NIH personel, who used the SCRC facilities on more than 1,400 occasions. The most popular application areas were image processing and molecular modeling, followed by sequence analysis and then statistical analysis. Macintosh® was by far the most requested platform, followed by Silicon Graphics and then PCs.

Researchers and their associates from virtually all NIH institutes, divisions, and centers have used the

consulting services and resources available through the SCRC. The greatest number of clients, total contacts and total hours of usage of the SCRC were from NCI, NIDDK and NINDS. However, among the heaviest *per capita* users were also NIDCD and NIAMS. The SCRC is currently used by 25 researchers per week, by appointment, averaging 30 new users each month.

Application Area Highlights

- *Image Processing*: The SCRC provides a variety of image acquisition and processing capabilities in its Image Technology Center (ITC). The ITC's primary focus has been on supporting the use of the NIH IMAGE program, developed by NIMH's Wayne Rasband. Popular application areas include the use of IMAGE for electrophoretic gel scanning, area measurements, spatial comparison, and image enhancement. More information about the SCRC Image Technology Center follows this section.
- *Molecular Modeling*: The SCRC provides access to a variety of molecular modeling software, including three of the most popular multipurpose molecular modeling programs: Quanta, Sybyl, and Insight II. The software is operational on two Silicon Graphics Indigo workstations, equipped with high performance graphics capabilities. SCRC molecular modeling software can be used to assist in the study of a wide range of biological molecules, including proteins, peptides, nucleic acids, polysaccharides, lipids and drugs. Applications include molecular structure prediction, protein structure-function relationships and drug design. The SCRC has added convenient access to large amounts of molecular modeling information, including interactive 3-D protein structure viewing, via MOSAIC. A Macintosh®-based 3-D protein structure viewing program, MacImdad, has been site-licensed for NIH use, and is distributed by the SCRC.
- *Sequence Analysis*: Because no single sequence analysis software adequately addresses every research need, a number of different sequence analysis programs are installed in the SCRC. Their emphasis ranges from selection and optimization of primers for Polymerase Chain Reaction (PCR) and sequencing reactions, to software for reading sequencing gels, to versatile programs designed to cover many major analytical needs in the laboratory. As new versions of sequence analysis software become available, they are evaluated and/or acquired and/or upgraded for use by the research community. Virtually every computational analysis feature offered either commercially or via freeware is currently ready for use in the SCRC.
- *Statistical Analysis:* There are a wide variety of statistical software packages installed in the SCRC for the Macintosh® and PC platforms. The different types of software products include: (1) the more widely-used, large and comprehensive packages; (2) the newer, less expensive and comparatively easy-to-learn packages (particularly for the PC Windows® environment); (3) the newer, visually-driven packages designed for exploratory data analysis; and (4) the high-quality products used for very specialized statistical applications and analyses. The type of analyses available include regression, exact and asymptotic logistic regression, analysis of variance and covariance, categorical data analysis, general linear models, non-linear curve fitting, 3D rotating plots, exact non-parametric inference, survey data analysis and survival analysis.

SCRC Outreach

As part of the SCRC outreach, Dale Graham visits NIH labs to give them tips on effective use of computing in the research environment. Graham also provides consultation to NIH researchers on a variety of computing topics, including sequence analysis programs on the Macintosh®, pedigree drawing in support of genetic analysis for the PC and Mac, Grateful Med, GOPHER, and MOSAIC.

To facilitate communication among DCRT and NIH researchers, the SCRC set up the Computational Sequence Analysis Resource (CSAR) list server, which is open to all NIH employees. CSAR is a convenient way to broadcast questions and informational replies via e-mail to the 60 current CSAR list members. Questions and answers are also archived to a CSAR GOPHER site at NIH.

In addition, SCRC is utilizing new tools that have become available for network access, such as GOPHER and MOSAIC. The SCRC has set up a World Wide Web (WWW) server to distribute SCRC related information to the NIH community, including many hypertext manuals and documents relating to the use of scientific software. To facilitate NIH access, a SCRC "Home Page" was created within the already established NIH Home Page, and it may be viewed with a WWW browser, such as MOSAIC.

By year's end, SCRC was actively exploring software programs and arrangements to create cross campus network access to a variety of commercial

SCRC programs for evaluation and use in short term projects.

SCRC Future

After the end of the second year of operation, the SCRC continues to readjust and realign its scope and directives to reflect the strategic plans of DCRT and NIH, and in response to feedback from our staff and users. We are in a continual phase of upgrading SCRC software resources to remain current and to deploy new computational tools as they become available.

Frequently, scientists arrive in groups of three or more to examine SCRC software or consult with SCRC staff. We hope to be able to add a combination demonstration and work area that is accessible to groups of three or more users simultaneously to accommodate this requirement.

Over the next 12 to 24 months, three areas of priority will be explored. We have a very strong interest in strengthening and improving network access to SCRC software from anywhere on campus. Network access will be a high priority in the coming year. Secondly, our two years of experience have clearly demonstrated a widespread NIH need for a scientific and technical graphics facility that the SCRC cannot address with current resources. We will explore better ways to address that need, perhaps with additional resources and/or in collaboration with others. Thirdly, the SCRC will continue to expand its outreach role. We will provide timely information and documents on our WWW SCRC server and facilitate other means of electronic communication and exchange, such as CSAR-L list servers and GOPHERs.

The SCRC Image Technology Center

David Chow

The Image Technology Center (ITC), in collaboration with the Scientific Computing Resource Center, opened in July 1993. It provides NIH with shared-use image processing computer systems, where NIH researchers are able to focus on scientific and medical imaging applications. DCRT imaging consultants provide guidance in the selection and effective use of advanced imaging hardware and software.

In addition to the high end Macintosh®, UNIX® workstations, and VMS® systems, the ITC has high-resolution image acquisition tools (scanners, digital cameras, film scanners), graphics workstations, and high-quality output devices.

Any NIH employee may seek assistance or use the scientific applications for evaluation or for short–term projects. The goal of the ITC is to make available different types of scientific computing solutions for imaging, so that researchers can make informed decisions about which of these they should incorporate as a resource in their own lab.

The First Year of Operation

In FY94, researchers and clinicians consulted with the ITC Staff on 776 occasions, and 1,632 ITC staff hours were devoted to studying imaging problems of mutual interest. The more popular application areas include electrophoretic gel scanning, area measurements, spatial comparison, image enhancement, segmentation, cell counting, 3-D reconstruction, collages, stack manipulation, superposition, densitometry and video microscopy. High demand exists for the Macintosh® imaging station running NIH IMAGE.

Researchers and their associates from most of the NIH ICDs have used the center's consulting services and resources. Even though the ITC opened only a year ago, statistics show that it has an average usage of 12 researchers/clinicians per week and the demand for services is increasing.

ITC Training Classes

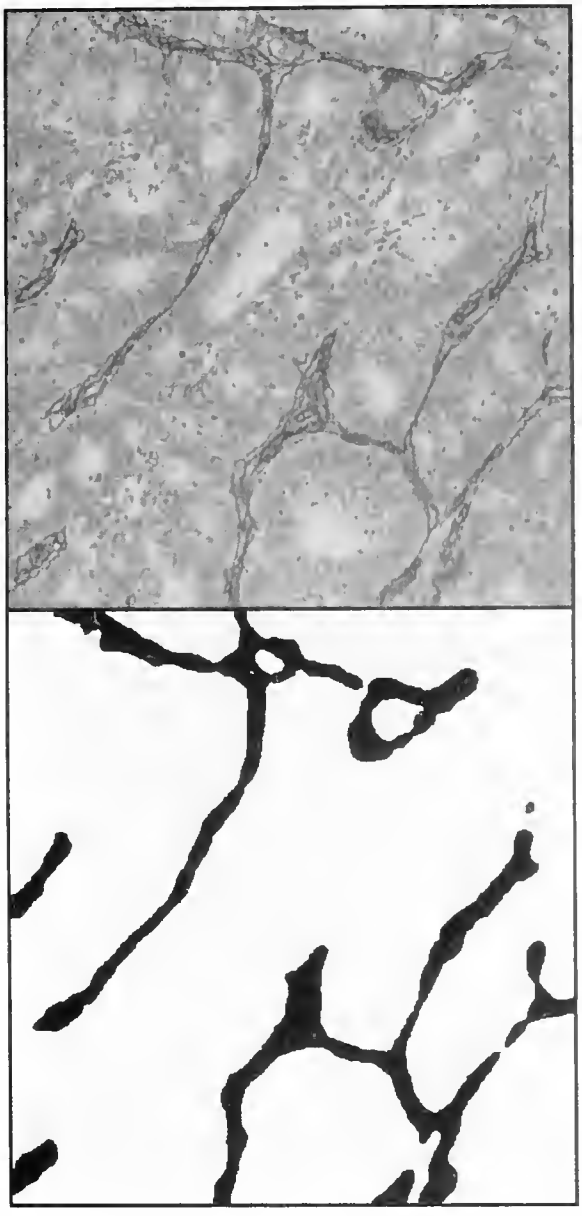
A total of ten image processing seminars were provided as part of the DCRT Training Program:

- Introduction to Image Processing (Benes Trus, Computational Bioscience and Engineering Laboratory, DCRT)
- Image Processing on the Macintosh® (Wayne S. Rasband, Intramural Research Program, NIMH)
- Using IMAGE 1.53 for Densitometric Analysis of 1-D Gels (David Chow, Distributed Systems Branch, DCRT)
- Selected Topics in Analyzing 1-D Gels (David Chow, Distributed Systems Branch, DCRT)
- Image Reconstruction via the Maximum Entropy Method (Peter J. Steinbach, Laboratory of Structural Biology, DCRT)
- Sessions on "1-D Gels" held May 31 and July 14 featured demonstrations in the ITC, a tour of the SCRC, a lecture and lab exercises using Quadra 800 Macintosh®.

Collaborative Projects

A large number of research projects are being investigated in the Image Technology Center. Techniques used range from very elementary to advanced.

NIH IMAGE is a public domain image processing and analysis program written by Wayne Rasband (NIMH) for the Macintosh®. This program has a number of important image processing functions and is being continually improved. The ITC tracks and participates in the development of the program, macros, and documentation, in addition to providing support, training and assistance in using NIH IMAGE.



Original image of the epithelial basement membrane and the mask produced with image processing.

Detection of basement membrane collagen-IV in colorectal tumors is a difficult problem being studied with R. E. Hewitt (NCI/DCBDC). Studies suggest that increased proteolysis accounts for the epithelial basement membrane (EBM) breaks that are common in carcinomas (see Fig. on this page). Deficient EBM production can be an important issue to investigate, and computer image processing techniques are being tested in an attempt to produce an efficient way of analyzing vascular basement membranes.

Displaying data is very important in the presentation of research, and is often useful in understanding or visualizing scientific phenomena. J. Chang (NHLBI/IR) used SpyGlass Dicer® in the ITC to produce 3-D reconstructions. Images of very thin (90 nm) slices of human monocyte were captured in NIH IMAGE using a Cohu® camera. Vacuoles in cultured white blood cells were imaged with Dicer®.

Imaging often involves the effective combination of hardware and software, utilized in a specific order to yield desired results. Q. Deng (NIA/LN) needed to generate superimposed images and group them in collages. NIH IMAGE was used to isolate the subjects in a technique involving "Gauss convolution," density slicing, and binary conversion. IPLAB® was used to form the superimposed image. Collages were formed using Photoshop and the final product was generated on a Phaser IISDX printer .

M. Levav (NIMH/LPP) conducted a neurological study involving epidemiology and intervention. A sample of 194 children ages 9-13 years was assessed with neuropsychological tests, CTscans and EEGs. CTscan data was analyzed in the ITC to correlate results in the neuropsychological assessments and the brain structures. The effects of malnutrition, iodine levels, goiter, and parasitic infection (neurocysticecosis) on brain volume and density were evaluated.

1-D Gel Analysis in the ITC generally involves the comparison of molecular weight based on bands that are optically measured (Fig. page 98). Consultation, group sessions, and seminars accommodate the large interest in this area. Rasband's improved "Gel Plotting Macro" was immediately supported by the ITC. In addition, staff produced a 23-page manual, "Using IMAGE for Densitometric Analysis of 1-D Gels," and 150 copies have been distributed.

ITC Future

Over the next 12 to 24 months, we plan to expand the ITC facility to provide additional computing power for faster processing of complex images, hardware for enhanced image acquisition, software for greater function and ease of use in image processing, enhanced network connectivity for more efficient image transfer, removable mass storage devices to increase

compatibility, and upgraded software (such as 680x0 to PowerPC® versions) for greater productivity.

Plans for expanded services include an information resource library, more ITC sponsored training and seminars, coordination of beta–testing and "seed" programs for new imaging products, development of a GOPHER/MOSAIC imaging resource, and user group gatherings.

The BioInformatics and Molecular Analysis Section

John Powell

The BioInformatics and Molecular Analysis Section (BIMAS) is dedicated to providing the NIH community with resources for high-level management and analysis of data in the field of molecular biology. This includes providing expertise in software development, applying computational analysis techniques to biological data and providing tools for accessing and displaying large amounts of genomic data from a variety of distributed databases.

A crucial component in the recent major advances in genomics research has been the uniting of advances in biology with those in computers, informatics, and networking. As genome sequencing throughput increases, the technological burden will shift increasingly to analysis and informatics. BIMAS was established to ensure that as this process advances, the necessary computational tools and resources are available to the NIH community.

BIMAS staff have undertaken a wide range of initiatives. BIMAS operates an ABI Inherit® sequence analysis system as a shared resource for the NIH community. BIMAS also administers a SYBASE database server for the development and testing of genomic databases. Staff members evaluate and support sequence analysis and related software via the ALW platform.

Collaborations and Support in BioInformatics Areas

Discovery of Novel Human Genes by Automated Sequencing of cDNA Libraries

J.I. Powell with L. Staudt, M.D., Ph.D. (NCI/DCBDC); R.C. Taylor (DCRT/DSB/BIMAS)

A collaboration is under way with L. Staudt, NCI, on the discovery of novel human lymphoid-specific genes by automated DNA sequencing of subtracted cDNA libraries. Software tools developed by DCRT are used to process and place the data into a SYBASE relational database system. These include: prescreening cDNA sequence against a local database; automating searching against the nonredundant databases on the NCBI network BLAST server; providing for display of the results; and allowing user interaction to select information to be placed into the SYBASE database.

Work is under way to provide software to perform complex motif pattern matching analyses, such as searches for nuclear localization signals, on the cDNA sequences. This software, based on Genobase and its associated toolkit, will permit automated incorporation of results into the SYBASE database, with a graphical user interface for input and editing of search parameters.

To date, thousands of cDNA sequences have been analyzed, yielding homologies to a variety of proteins including transcriptional regulators, signal transduction proteins and membrane receptors. Work is in progress to expand the scope of the database to include laboratory management information and data from other sources such as northern blots.

Support of Sequence Analysis and Related Software via ALW Platforms

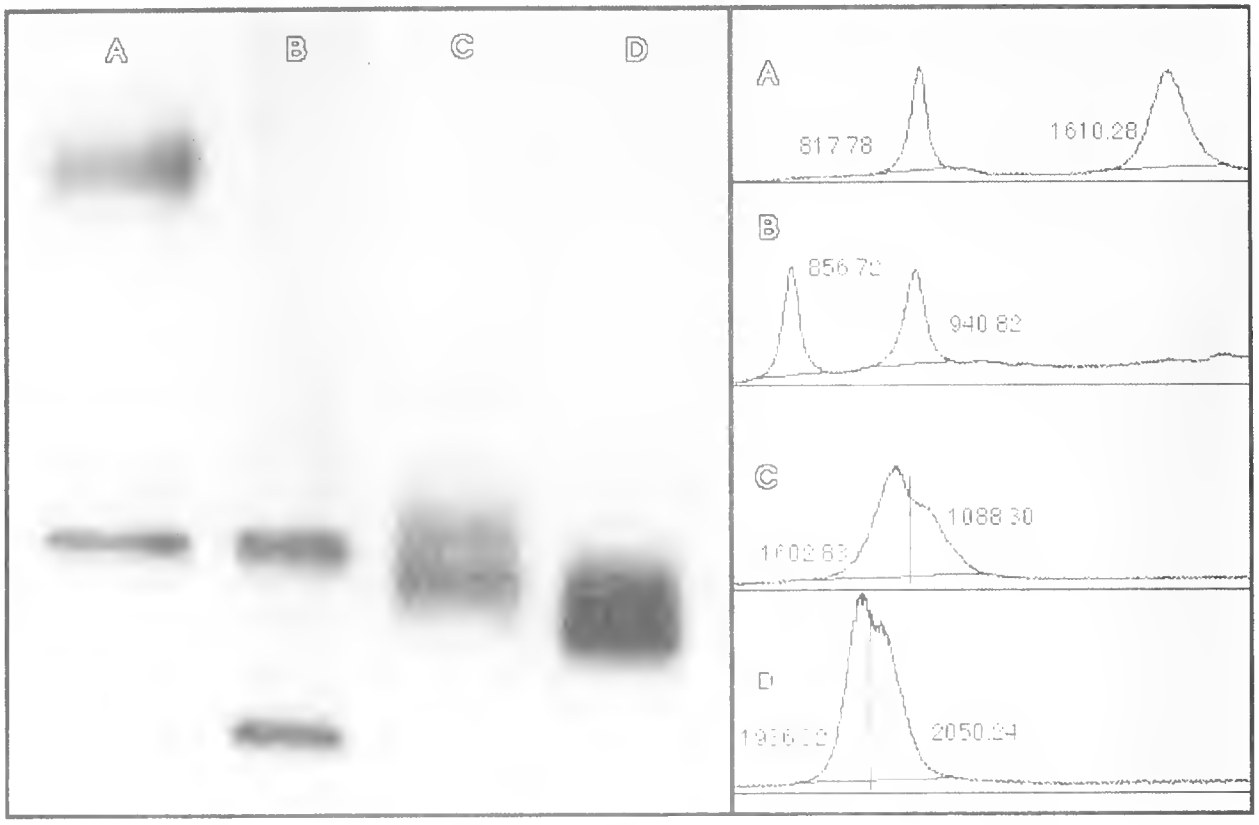
J.I. Powell

BIMAS makes available and supports a number of genomic software packages under DCRT's Andrew File System. These packages fall into the following categories: comprehensive genomic sequence analysis suites, such as the Genetics Computer Group (GCG, based on the Wisconsin Package), and GDE (the Genetic Data Environment); genetic mapping tools and utilities, such as CRI-MAP, QUICKMAP and SIGMA; sequence similarity data base search tools, such as Fasta, Local BLAST and Network BLAST; sequence data base access tools, such as Network Entrez and the sequence analysis tool Signal Scan.

ABI Inherit® Sequence Analysis Resource

J.I. Powell

DCRT is making available an Applied Biosystems, Inc. Inherit® system as a shared resource to the NIH intramural research community. This system employs a client/server architecture using an Apple Macintosh® computer as the client platform. Scientists can purchase client software from ABI and access the Inherit® system over the NIH network. To speed results, Inherit® makes use of highly specialized hardware. The Fast Data Finder™ (FDF) parallel processor can perform parallel pattern matching searches through large databases at a rate of over 15 million characters per second. This speed permits completion in



Electrophoretic 1-dimensional gel and the profile plots of the measured optical densities (see page 96).

hours of tasks that often require days using powerful UNIX® workstations.

The system is best suited to: (1) assembly of medium to large sequences; (2) searching gene and protein databases for sequence homologies; and (3) rapid searches for genetic motifs, such as regulatory elements. A pattern description language integral to the system permits construction of very complex queries. DCRT has provided considerable feedback to ABI to improve the client user interface, as well as exploring the possibility of porting client software to additional platforms, such as UNIX® workstations or the NIH CONVEX/SGI server.

This project highlights the potential of the NIH network to bring powerful and sophisticated resources to the scientist's benchtop through desktop computers.

A Critical Evaluation and Comparison of Computerized Sequence Analysis Programs

J.I. Powell
with M. Miller, Ph.D. (NCI/LEC)

In collaboration with M. Miller, NCI, a critical, assembly and analysis packages was undertaken. A fundamental problem in contemporary molecular biology is the determination and interpretation of DNA sequences. Due to limitations of current sequencing technology, sequence determination entails the piecing together of short, overlapping sequence fragments into a single, long contiguous sequence. A number of commercial computer programs have been marketed to automate this process. While reviews of individual packages have been published, this is the first known study to critically compare the accuracy of assembly by these programs.

Eleven programs were selected, primarily on the basis of their availability on the NIH campus. Sequence data is not random, but contains ordered repeated sequences. Likewise, errors in sequencing determinations are not randomly distributed. In order to provide a controlled and realistic dataset for measuring performance and accuracy, a known sequence, the rat multidrug resistance gene (RATMDRM, 5254 base pairs, accession number M62425) was split into 58 random overlapping fragments of 200 to 400 base pairs in length. These were then randomly seeded with 0 to 15% error based on the error distribution of the fragments originally used to determine the sequence.

Errors were in the form of miscalled bases, deleted bases or added bases.

Based on accuracy, the programs tested fell into three general groups. In order to rule out conditions unique to the chosen test sequence, four other sequences between 4,500 and 4,600 base pairs were used to repeat the tests. With one exception, the error rates were comparable to those encountered using RATMDRM. Additionally, some programs were tested with different permutations of RATMDRM to ascertain their capacity to properly assemble the sequence regardless of the order of input of the fragments. Ease of editing the assembled sequences was also compared. Results of this study were accepted for publication by the Journal of Biological Computation.

Publications

Miller MJ, Powell, JI. A critical evaluation and comparison of computerized sequence analysis programs. J Biol Computation (in press).

Genobase, An Object-Oriented Genomic Database Accessible through the WWW/MOSAIC

R.C. Taylor

Work is in progress to develop and provide an integrated framework for computational support of research in comparative DNA/protein sequence analysis and related areas across multiple genomes/species. The logic programming language PROLOG is used throughout this project, permitting data of disparate types to be combined rapidly and effectively, and permitting complex queries from the integrated data. Toolset development was performed in close collaboration with R. Overbeek and R. Hagstrom of Argonne National Laboratory.

Current work focuses on the combining of large volumes of data from multiple sources, such as GenBank, EMBL, Prosite, SwissProt and others including metabolic data from Overbeek's Russian collaborators. This will form an integrated database with DNA and protein sequence, motif, metabolic pathway and other data for multiple genomes, resulting in a unique resource. The work incorporates analysis of genomic organization and genetic regulation of metabolic pathways. This database and associated tools will permit answers to queries difficult or impossible to satisfy using the currently available biological databases.

An experimental implementation is presently available, and is supporting intramural research at NIH by L. Staudt, NCI. This database is also the underlying data repository for a World Wide Web (WWW) hypertext browser implemented by R. Taylor and A. Ginsburg, DCRT/BIMAS. It is expected that this WWW service will provide a unique resource to the biomedical research community over the Internet, employing simple and widely available end-user client tools such as NCSA MOSAIC for access. This will supplement present WWW servers at NCBI and at EMBL in Europe, providing unique services unavailable to date.

An additional component is the inclusion of data from the new NCBI MacroMolecular structure Database (MMDB), which contains three-dimensional crystallographic structure data with ties to and from sequence data. Efforts are presently under way to add three-dimensional structure-viewing capabilities to the Genobase toolkit.

Support of NINDS Neurogenetics Core Sequencing Facility

J.I. Powell
with R.C. Taylor (DCRT/DSB/BIMAS); L. Goldfarb, M.D., M. Dubnick (NINDS/DIR)

In an ongoing collaboration with Lev Goldfarb and Mark Dubnick, NINDS, a database system is being developed for the study of inherited neurological disorders. The system combines neurological examination data with pedigree information in order to facilitate statistical and epidemiological studies of genetic neurological diseases, such as Machado-Joseph's disease, autosomal-dominant cerebellar ataxia, and familial Alzheimer's disease. The data for such studies consist of one or more neurological examination reports on various patients in a kinship and information concerning the relatedness of family members, as well as the disease state observed or inferred for other kinship members not examined. The database is implemented using a SYBASE relational database, which contains the neurological examination data and tables coding the relationships among family members. The relational database also contains index information relating individuals with the examination records and family histories in which they are mentioned. Alone, the relational database would allow a variety of statistical studies of the disease data.

In order to expand the utility of the system, and to allow more complex genetic queries, a sophisticated Prolog-based query interface is being developed. Prolog is an implementation of the processes involved in first-order predicate calculus. A Prolog program is a set of clauses. Each clause expresses either a fact or a rule, and a Prolog program draws conclusions and makes inferences from the knowledge or information contained in the clauses. An inference engine built into Prolog

performs searches and pattern matching automatically. When a query is made, the inference engine is invoked. The engine then uses a depth-first search strategy combined with backtracking to examine the facts and rules in the database (which is formed out of clauses of the program itself) for evidence to answer the query. ProDBI, a Quintus Prolog/SYBASE Interface for SUN® 4, allows SYBASE tables to be accessed from Prolog as though they existed within Prolog's environment as unit ground clauses (facts). The interface module will enable our Prolog engine to interactively treat each row (record) of the tables of our SYBASE database as a native Prolog clause, with each column (field) treated as an argument of the clause. Prolog's capability to make more complex queries in more natural language than Sybase SQL will be especially useful, since these data will be accessed by people with clinical and epidemiological backgrounds.

Furthermore, software will be written to generate a description of any desired pedigree for input into a commercially or academically-available pedigree drawing graphics package, in order to display arbitrary pedigrees when desired. It is hoped that this system will prove useful for the study of a variety of genetic neurological diseases, in addition to those studied by the authors. The examination data are designed to be comprehensive enough to accommodate a complete generic neurological examination.

In addition, DCRT continues to provide technical and administrative support for computer systems used for DNA and protein sequence analysis in the NINDS Core Sequencing Facility. DCRT personnel recently participated in installation and configuration of an upgrade replacing several older computers with modern distributed workstations. These collaborations will continue as the computing requirements of this facility evolve.

New Technology Tracking/Evaluation
J.I. Powell

DCRT has a responsibility to keep abreast of the latest advances in bioinformatics and genomic/sequence analysis in order to continue to provide strong and effective support to the NIH intramural research community. Access to sequence data is rapidly becoming the central element in molecular biology research as well as playing a key role in new and innovative approaches in pathophysiology and therapeutics.

In addition to evaluating and supporting new commercial technologies and products such as the Inherit® sequence analysis engine, DCRT personnel regularly attend scientific meetings at Cold Spring Harbor and other sites to stay abreast of the scientific progress in this vital field. DCRT personnel maintain collaborations and contacts with personnel at NCHGR, NCBI, DOE and those in other organizations who set standards and provide central data repositories.

DCRT staff continue to develop innovative and effective interfaces to this data to make it more accessible and useful to NIH research scientists. A key example is the DCRT experimental Genobase server, which promises to provide integration of many forms of molecular biology data in a single resource, permitting novel uses.

Support in Areas of Laboratory Automation/Analysis/Archival Storage
J.I. Powell

DCRT provides a wide variety of services in support of computer automation of laboratory bioinformatic and genomic data. These services include support of computer systems administration, installation, problem resolution, data security, data archiving and retrieval, and development of custom applications programs and user interfaces. Collaborations exist with NCI, NINDS, NCHGR and others to bring together the wealth of computing technology and expertise within DCRT and the biomedical scientific expertise of NIH intramural scientists and to form a synthesis empowering research scientists with the advantages of computer automation without the hurdles of jargon and complexity. Ultimately, the goal is for the scientist to address the computer in scientific terms instead of arcane and unwieldy programming languages. Bioinformatic data is the key to modern molecular biology and is expanding to other fields of biomedical research. The availability and accessibility of these data is crucial to progress.

By supporting bioinformatic software on the NIH ALW network, DCRT brings powerful tools and large storage capacity to bear on these issues. Development and installation of specialized hardware, such as the ABI Inherit®, and sophisticated software, such as the DCRT World Wide Web Genobase server, are bringing powerful and timely tools to the NIH intramural research community. It is expected that these tools will accelerate the discovery of disease etiologies and the development of new therapies. DCRT will continue to expand and improve on these collaborative efforts, as new opportunities develop for application of bioinformatic tools to contemporary scientific problems.

Flow Cytometry Collaborations and Support

Luther Barden

General Support for NIH-wide Flow Cytometry

The on-going seminar series "Topics in Analytical Cytology" hosted two sessions during the year under the auspices of the NIH Computer Training Program, with presentations by NIH, FDA and USUHS researchers in:

- Flow and Image Cytometry in B-cell Chronic Lymphocytic Leukemia
- Advanced Techniques in Quantitative Fluorescence Microscopy
- Studies of Drug and Carcinogen Efflux in Multi-drug Resistant Cells using Adherent Cell Laser Cytometry
- *In vivo* Confocal Microscopy of the Human Eye

During the coming year we expect to concentrate on evaluations of in-house and commercially available flow- and image-cytometry data analysis programs.

Cluster Analysis Program

The Cluster Analysis Program (CAP) has been ported from its originally designed VAX/VMS minicomputer and graphics terminal environment to a RISC OpenVMS Motif workstation platform, with some necessary changes to computational algorithms and data structures to take more complete advantage of the RISC architecture. During the coming year we will expand CAP's functionality in terms of implemented clustering algorithms and clustering strategy.

Laboratory Analysis Package

The Laboratory Analysis Package (LAP) was originally developed to run on SUN3 UNIX® workstations as a general-purpose tool for both interactive and batch processing of laboratory data. LAP is currently implemented in C++ version 2.1. and has been ported to SUN4, VMS® (VAX and Alpha), and Convex architectures. It is used extensively by two laboratories in NIDDK and numerous Flow Cytometry sites at NIH.

LAP can perform a wide range of data manipulations on vector data, x-y paired data and matrix data, using either a command or expression syntax. Customized command procedures can be saved in files and added to the LAP command set. Results may be viewed as line graphs, scatter graphs, bar graphs, perspective views or contours in color or monochrome viewports within an X window or on Tektronix 4010 and 4107 compatible terminals. Publication quality plots may be produced in several formats, including encapsulated postscript and HP-GL.

This year, the User's Guide and Reference Manual were updated, porting to OpenVMS AXP was undertaken, and an FTPable distribution was made publicly available. In addition, we offered a LAP training course for Flow Cytometry users. Future effort will be limited to maintaining the program.

Support for the OpenVMS Operating System

Currently there are approximately 100 VAX/VMS systems at NIH, which provide computing services to nearly 1,000 users. In recognition of the importance of these systems to the NIH scientific community, DSB continues to provide a stable level of support for OpenVMS on both VAX and Alpha platforms. Both new hardware capacities and enhanced software capabilities will ensure our ability to support members of our user community as they migrate to expanded capabilities as well.

Hardware and software support resources are provided for both VAX/VMS and AXP/VMS computing operations throughout the NIH. Software tools include language compilers, database systems, consultation services, and connectivity products. The three-member VMS cluster has been enhanced by upgrading the MicroVAX II to a VAX 4000-100 with 80 Mbytes of memory and an additional 7 Gbytes of disk space and by the addition of a DEC AXP workstation. Supported software includes: FORTRAN, Pascal, C, and C++ compilers; several code development support utilities; network diagnostic utilities; distributed file and queuing service software; Rdb database development tools; screen and hardcopy graphics support libraries; and a large library of user-contributed software. We are now adding personal computer file and print service support via the Pathworks product suite. Consultative services are still furnished primarily through the contractor-operated "VMS Hotline" as an adjunct to consultation provided by DSB staff.

The major addition to the VMS support effort has been the acquisition of a DEC 3000-400 AXP RISC-based workstation. This system, which is equipped with several compilers and code development and migration tools, forms the heart of a VAX-to-AXP VMS migration assistance resource, and will enhance our ability to assist our user community during migration of their applications to the high performance computing solutions of tomorrow.

Laboratory and Clinical Applications Section

Kenneth Kempner

The Laboratory and Clinical Applications Section (LCAS) provides collaboration, consultation, and support within NIH Laboratory and Clinical environments. LCAS capabilities cover both hardware and software aspects of the implementation of minicomputers, microcomputers, workstations and computer networks.

Laboratory Projects that involve LCAS staff typically concern data collection, database, control, and interfacing efforts for many types of analytical instrumentation. Clinical projects currently undertaken by LCAS staff members range from 2-dimensional signal analysis in the time and frequency domains, through 2-dimensional image processing, to 3-dimensional image reconstruction. In addition, LCAS engineers have expertise in computerizing diagnostic electrocardiography, ECG arrhythmia monitoring, multiparameter physiologic monitoring, and echocardiography.

Many LCAS projects are conducted within the Clinical Center (CC), and many of these involve direct assistance to the mission of a CC department. As a result, the LCAS provides support to the laboratory and clinical activities of all institutes that maintain facilities within the CC. The long-term support of the ECG Data Management System, which provides automated diagnosis, storage and retrieval of 12-lead diagnostic ECGs obtained from all CC patients, is an example of a clinical application that directly serves all NIH Institutes.

Electronic radiology, multimedia imaging, and large-scale image archiving are important emerging technologies that have the interest and attention of the LCAS. Potentially, these areas will be of major benefit to the NIH community, and prototype projects within these three spheres are currently in progress or being planned.

Real-Time Gamma Camera Image Correction

W.R. Gandler
with M.V. Green, and J. Seidel, Ph.D. (CC/NMD); and K.M. Kempner (DCRT/DSB)

The Nuclear Medicine Department (NMD) of the Clinical Center has developed a small Field-Of-View (FOV) gamma camera, which has great promise for practical, high-resolution imaging of small animals. The system is based on a novel position-sensitive photomultiplier tube (PMT), instead of the multiple non position-sensitive PMTs used in standard, large FOV gamma cameras. Unfortunately, the position-sensitive PMT does not possess either a linear voltage analog of event position, or a uniform energy response across the tube face.

Three coupled 386/486 processors comprise a Multibus II Image Correction System that allows first-order, geometric and energy corrections to be performed sequentially, in real-time on data from the small FOV gamma camera. The Image Correction System either acts as a stand-alone, two-processor data acquisition system for the small FOV gamma camera, or it is interposed between this camera and a commercial Nuclear Mac™ Analog Acquisition System, and is used as a three-processor system, dynamically correcting the data transmitted from this camera to the Nuclear Mac. The three processors are dedicated to input (analog-to-digital conversion), computation (geometric, energy and motion correction), and output (digital-to-analog conversion or digital transmission), respectively. Software for system control, data acquisition, corrected and uncorrected image display, and data/image transmission has been developed. All geometric and energy correction software has been completed. It is possible to acquire up to ten simultaneous inputs via the high-speed analog-to-digital converter module.

In addition, unique software has been developed to provide the real-time, second-order manipulations necessary to allow off-line reconstruction of PET images from only two position-sensitive photomultiplier tubes, in a single-slice configuration. These manipulations include the determination of coincidence lines for all positron annihilations, the sorting of coincidence lines into projections, one for each angular position around the object, and the production of an image-like sinogram from these multiple projections. Software that displays the resulting two-dimensional X versus Y sinogram arrays or three-dimensional X versus Y versus Energy arrays has been implemented on both the Multibus II and the Macintosh® platforms.

Previous collaborative efforts between the NMD and our laboratory have, independently, demonstrated the need for methods to correct for motion artifact during planar gamma camera studies of the brain. A Brain Position Measurement System (BPMS) was developed, incorporating a Polhemus position/orientation measurement subsystem that produces the necessary corrections. Motion correction software has been written for the Multibus II Image Correction System,

based upon the integration of real-time data from the independent PC-based BPMS.

In the future, sinograms will be produced from data acquired in a multiple-slice PET configuration. Two-dimensional reconstruction software has been completed by Nuclear Medicine Department Staff, and three-dimensional reconstruction software is under development. A stand-alone data acquisition system, residing entirely on the Macintosh® platform, is in the preliminary stages of development

Work is also beginning on the implementation of new algorithms for image acquisition in PET scanner mode that are suitable for use in imaging the human breast. This joint effort, involving DCRT, CC, NCI, and BEIP personnel, will explore new approaches to the early detection of breast cancer.

Presentations

Seidel J, Gandler WR, Green MV. A very high-resolution single-slice small animal PET scanner based on direct detection of coincidence line endpoints (Abstract). J Nucl Med 1994;35:40P. Presented by J. Seidel at the 41st Annual Meeting of the Society of Nuclear Medicine, Orlando, 1994.

Green MV, Seidel J, Gandler WR. A small animal system capable of PET, SPECT and planar imaging (Abstract). J Nucl Med 1994;35:61P. Presented by M. Green at the 41st Annual Meeting of the Society of Nuclear Medicine, Orlando, 1994.

Seidel J, Green MV, Gandler WR. Estimated spatial resolution of a single slice, small animal PET scanner using position-sensitive photomultiplier tubes for event pair location (Abstract). Proceedings of the 29th Annual Meeting of the Association for the Advancement of Medical Instrumentation, Washington, DC, 1994, p. 85. Presented by J. Seidel.

Image Management and Communications System (IMACS)

K.M. Kempner
with H.G. Ostrow (DCRT/NSB); T.L. Lewis, M.D. (CC/DIR); E.E. Tucker, M.D. (NHLBI/CB); J. Doppman (CC/DRD); P.G. Okunieff, M.D., F. Sullivan, M.D. (NCI/ROB); and J.F. Fessler (NCRR/BEIP)

Medical images are an important component of the medical record generated during a patient's hospital stay or clinic visit. Unfortunately, these images represent a data source that is difficult to manage because of the extremely large size of the datasets involved. The NIH Clinical Center (CC), like most university and research hospitals, is attempting to solve the problem of consolidating medical images with the conventional alphanumeric medical record data in the Medical Information System (MIS) to more completely realize the goal of a comprehensive electronic medical record. Toward this end, DCRT, CC, and NCI are collaborating to develop a series of demonstration projects that explore image integration into the electronic medical record. The images of interest range in size from diagnostic electrocardiograms (16 KBytes) through tomographic scans (256 KBytes) to conventional film x-rays (4 MBytes).

A standard 12-lead diagnostic electrocardiogram (ECG) is automatically acquired, interpreted, and stored on magnetic disk, using a Hewlett Packard ECG Data Management System located in the CC. In order to transfer ECG diagnoses and the related waveforms from this minicomputer, a Hewlett Packard ECG Workstation was developed as a gateway between the two systems. Because ECG waveforms are essentially binary images (black waveforms on white background), and because the number of equivalent black pixels in such an image is extremely low (approximately 0.1 percent), ECG waveform data are more efficiently stored and transmitted through the gateway as files of about 17 KBytes, containing time-ordered lists of 10-bit ECG amplitudes, rather than as files of about 8.25 MBytes, containing 2.75K by 3K pixel images.

Chest x-rays are routinely obtained within the Diagnostic Radiology Department, and these images are appropriate for integration into the CC MIS, as well as for transmission to the outpatient clinic where the patient will be seen. In this application, we are currently using a Vision Ten Rita!® system, which contains a gray-scale sheet film digitizer, as an integral part of an image gateway. In addition, we have installed two Rita!-compatible image display systems. Utilizing the fiberoptic network installed within the CC, communication of medical images between the Radiology Department's Film Library and remote sites is now possible. The weekly NHLBI Cardiac Surgical Clinic was the first outpatient clinic to routinely use chest films transmitted over this Ethernet pathway.

Future Plans

Future plans include the connection of two General Electric 9800 CT scanners into the Vision Ten image transmission and display environment, in order to support the electronic imaging needs of the NCI Medicine Branch's outpatient clinic. This will be accomplished via ACR-NEMA-compatible, Ethernet-

based communication links to dedicated image servers, which will be added to the teleradiology network.

In addition, we are planning a prototype high-speed image communication network based on Asynchronous Transfer Mode (ATM) Switch technology. Initial plans specify the implementation of two ATM switches, with 16 ports each, which are capable of supporting multimedia communication at 155 Mbits/sec. One of the ATM switches will also serve as a testbed for the DCRT Network Systems Branch (NSB), in an evaluation of ATM technology as a backbone architecture for the dynamic interconnection of Clinical Center LANs.

This prototype network would initially support high-performance radiation therapy planning, which is a collaborative effort between DCRT's Computational Bioscience and Engineering Laboratory (CBEL) and the NCI Radiation Oncology Branch (ROB). CBEL's Intel iPSC/860 Supercomputer will be utilized to apply the power of parallel computing methods to the implementation of the computationally intensive calculations required for radiation therapy planning. A custom-designed Radiology Consultation WorkStation (RCWS) will be located in the NCI Radiation Oncology Branch (ROB), as well as in the same building as the CBEL Supercomputer.

Real-time consultation sessions between an ROB radiation oncologist or radiation therapy physicist and a CBEL domain expert should contribute to the rapid development of radiation therapy treatment plans. Eventually, a third RCWS will be located in the CC Diagnostic Radiology Department to allow three-way consultations that include a radiologist. The CT images to be transmitted via this system will be obtained from the same image servers that will ultimately provide images through the Rita! system to the NCI Medicine Branch.

The two ATM-compatible workstations that will be developed in support of the CBEL and ROB collaboration are to be optimized for radiology consultation purposes, in general, and radiation therapy planning, in particular. This Radiology Consultation WorkStation (RCWS) is a multimedia medical imaging workstation, which will be appropriately designed for use in an electronic radiology environment. Centered around a SUN® SPARCstation 20 workstation, each RCWS will include two high-resolution monochrome image display systems, designed to function as electronic view boxes for the presentation of "electronic films" in 14 X 17 inch format.

Brain Image Registration

K.M. Kempner
with M.V. Green (CC/NMD); J.F. Fessler (NCRR/BEIP)

The superposition and registration of differing tomographic views is a difficult problem for investigators attempting to correlate brain form (structure) derived from x-ray computed tomography (CT) images with brain function (metabolism) revealed by nuclear medicine Positron Emission Tomography (PET) images

Precise orientation of the subject's skull within the scanner's aperture is monitored and recorded with a Brain Position Measurement System (BPMS), consisting of a Polhemus position/orientation measurement subsystem connected to a PC. The development of two types of inexpensive, custom-molded oral appliance allows the Polhemus subsystem's sensor to be fixed to the subject's skull (see also page 95).

Next year we plan to upgrade the hardware with a faster microcomputer. Also, the BPMS will be interfaced to another device, the "Real-Time Gamma Camera Image Correction System." This will allow motion artifacts to be significantly reduced, if not totally eliminated, from brain images produced by the CC Nuclear Medicine Department's unique, Small-Field-Of-View Gamma Camera.

ECG Data Management System

K.M. Kempner
with E.E. Tucker, M.D. and J.E. Elson (NHLBI/CB); J.F. Fessler (NCRR/BEIP); and D. Zywicki (Systex, Inc.)

The NIH Clinical Center's (CC) heart station uses a computerized system for the analysis of the clinical Electrocardiogram (ECG). Hewlett Packard's ECG Data Management System (DMS) processes 12-lead ECGs from all patients within the CC. This system collects and processes ECG waveforms, measuring amplitudes, durations and intervals. It also provides a clinical diagnosis, allows editing of the diagnosis after physician review, stores the ECG waveforms and the diagnostic reports, and permits searching the database for patients who meet search criteria.

The medical diagnostic criteria are encoded as IF-THEN production rules contained in a Diagnostic Criteria Set. These rules were written using Hewlett Packard's Electrocardiogram Criteria Language (ECL), and the Diagnostic Criteria set may be modified by the user to tune existing criteria or to add new criteria. During the summer of FY94, a criteria adjustment was

performed to minimize the number of false positive and false negative statement occurrences, with the expectation of reducing the overread and editing efforts.

ECGs are transmitted digitally to the ECG DMS over 2400 baud dial-up telephone lines within the CC, from 14 ECG machines distributed throughout the facility. The ECG diagnostic reports, and ultimately the ECG waveforms, will be sent to the Clinical Center's Medical Information System (CC MIS) for display at any user terminal. A Hewlett-Packard ECG Workstation has been installed as a gateway between the ECG DMS and the CC MIS. Implementation of this bi-directional pathway was completed in August 1994, for the purpose of transmitting confirmed ECG diagnostic reports, which have been overread by a cardiologist and edited within the ECG DMS.

Current efforts involve the development of software to support the transmission of the immediately available, unconfirmed ECG diagnostic reports to the CC MIS. In addition, software is being developed to provide for the conversion of paper ECG Waveforms into appropriately-formatted files, which simulates analog-to-digital converter data. This will allow for convenient storage of this data within the ECG DMS. The ECG Waveform Conversion Subsystem will allow the CC to discard a large number of paper ECG records currently kept in a hardcopy file room because they were not originally acquired via direct telephone transmission to the ECG DMS.

Plans for FY95 also provide for the replacement of the current minicomputer-based, Hewlett Packard ECG DMS with an updated, microcomputer version. The new system will maintain the entire current database of over 125,000 waveforms and reports online, and will also be compatible with the CC MIS System and the ECG Waveform Conversion Subsystem.

3-D Flow Velocity Reconstruction from Color Doppler Ultrasound Images

D.R. Adam, Ph.D.
with K.M. Kempner (DCRT/DSB); M.A. Vivino (DCRT/CBEL); E.E. Tucker, M.D. (NHLBI/CB); T.J. DeGraba, M.D. (NINDS/SB); and M. Jones, M.D. (NHLBI/SLAMS)

Clinical color Doppler ultrasound technology is a popular, non-invasive, real-time, relatively inexpensive imaging modality, which currently allows the 2D visualization of blood flow within the heart and the vascular system. Doppler ultrasound flow velocity measurement is important for the determination of blood/oxygen supply to various organs, arterial wall shear stress, and blood-tissue gas exchange, as well as for the evaluation of myocardial and valvular function.

The current methodology for Doppler flow measurement was found to be misleading in some respects. Also, none of the present Doppler ultrasound systems measure the spatial position and orientation of the ultrasound transducer, and its relation to the flow direction. A procedure was developed that appears to lead to an accurate determination of flow velocity. Our methodology takes into account the spatial position and orientation of both the ultrasound transducer and the vessel being imaged. The ultimate goal is the quantification of vascular flow patterns, thus enhancing the usefulness of this important non-invasive diagnostic tool.

Initial studies concentrated on the structure and flow in the carotid artery, due to the simplifications that this geometry allows. Utilizing color Doppler ultrasound technology, in a clinical echocardiography laboratory, the carotid artery was imaged from several positions and orientations, producing data sets capable of generating three-dimensional reconstructions of this vessel's structure and flow profile. A carotid artery/neck phantom was designed and fabricated to allow for calibration and testing of both the position/orientation measurement subsystem and the Doppler flow velocity measurement subsystem, in a controlled environment.

A patent application, covering the basic algorithm for correcting the color flow velocity measurements, is in process. This project is otherwise currently inactive due to the return of Adam, who was a Senior Visiting Scientist, to his home institution. Work in this area is continuing at the Technion in Israel, under the direction of Adam.

Multimodality Research Image Processing System

Margaret Douglas
with P.J. Kalkowski (DCRT/DSB); R. Levin, Ph.D., D.G. Sobering, Ph.D. (NCRR); R. Carson, Ph.D. (CC/NMD); J. Frank, M.D., A. Polis (NINDS)

DCRT has continued to support the efforts of LDRR and NCRR, and several ICDs to develop a common, shared approach to medical image processing. DCRT staff have coordinated the meetings of representatives from the ICDs, have assisted in designing software, have monitored and evaluated software developed by a vendor, and have assisted in the management of the project. The vendor has provided extensive software, involving 1.5 million lines of code.

DCRT has collaborated with NCRR and NMD (CC) to test this code, and has identified and prioritized design problems. Under the aggressive management by DCRT, the vendor has substantially improved the code and reduced the number of problems ("bugs"). Due to limited resources and the retirement and reassignment of two DCRT staff, the future of this project will be determined by LDRR and the collaborating ICDs.

Biostatistical Studies and Consulting

James Malley, Ph.D.

Malley provides selected mathematical and statistical consulting as part of the staff of the Office of the Chief, DSB. This year saw the continued growth of the use of adjunct staff dedicated to providing data analysis and experimental design support for researchers in a number of institutes in the NIH Clinical Center, many of whom do not have easy (or any) access to the consulting skills and knowledge of other statisticians at NIH. This support group, organized and coordinated by Malley, now has five members, and this year over 400 client contacts were handled, from nearly a dozen ICDs.

Malley continued to develop and teach data analysis workshops and seminars throughout the year. A five part series on recurrent data analysis was presented, in addition to a workshop/seminar on Scientific Data Analysis, entitled "Methods and a workshop/seminar on Scientific Data Analysis: NIH resources."

A long-term collaboration began this year with leading-edge statistical software developers. One result was the invention of an interactive and highly efficient visual nonlinear curve fitting scheme, which will now be used as a front-end routine in the exploratory data analysis system Datadesk®. Work is continuing with Datadesk® to construct self-contained multimedia statistical software that can serve as both a data-driven, real-world statistical analysis instructional environment, and also as an efficient, highly-interactive data analysis system.

Similar to this has been an implementation of the analysis scripting language of HiQ©, an advanced numerical analysis and modeling system (on the Mac and UNIX® platforms). A linked series of estimation routines was identified and implemented to perform an advanced maximum likelihood estimation method for rapid determination of whether a given biomedical data set may be a two-component mixture of normally distributed subgroups, rather than a single normally distributed population. The procedure was applied to studying the possibility that distinct subgroups exist in children with learning and attention deficits, using data collected at NIMH by Norman Rosenthal (CBP).

Malley continued his study of the practical and theoretical consequences of statistical data analysis methods modified to deal with biomedical data that is dominated by quantum–mechanical noise. On the practical side, a number of existing biomedical technologies (such as the scanning tunneling microscope, the atomic force microscope, laser cooling and trapping of single atoms and molecules, and ultrafast pump-probe laser spectroscopy) are already operating within this domain.

Theoretical work includes a continued investigation of new problems and opportunities available for optimally efficient statistical data analysis methods for quantum events. The research has stirred interest in the physics and mathematical statistical communities, and results have been presented at invited lecture series at Harvard and M.I.T. and at the University of Aarhus, Denmark, where Malley gave lectures to the statistics department staff and continued his joint work with O.E. Barndorff-Nielsen.

In November, 1993, the peer-reviewed journal *Statistical Science* published "Quantum Statistical Inference," an invited research and review article (co-author: John Hornstein, Naval Research Laboratory, Washington, DC), covering the subject in theoretical detail and giving practical examples. A more refined theoretical analysis of quantum statistical inference methods has been completed, and will be submitted to "Bernoulli," an international mathematical statistics journal. Some of this work was done jointly with Barndorff-Nielsen.

In another research direction, multiple review cycles were completed for the monograph "Statistical Applications of Jordan Algebras." It was published in August 1994 by Springer-Verlag. This study is now the focus of some attention: Malley has been invited to attend an international work group (Oberwolfach Conference; Freiberg, Germany; July 1995) on the use of algebraic methods in multivariate data analysis.

Publications

Knebel A, Janson-Bjerklie S, Malley J, Wilson A, Marini J. Comparison of breathing comfort during weaning with two ventilatory modes. Am J Resp Crit Care Med 1994;149:14-18.

Rosenthal N, Brown C, Oren D, Galetto G, Schwartz P, Malley J. Effects of light on T-cells in HIV-infected subjects are not dependent on history of Seasonal Affective Disorder. Photochem Photobiol 1994;59:314-319.

Dichek H L, Nieman LK, Oldfield EH, Pass HI, Malley J D, Cutler GB. A comparison of the standard high dose dexamethasone suppression test and the overnight 8-mg dexamethasone suppression test for the differential diagnosis of adrenocorticotropin-dependent Cushing's Syndrome. J Clin Endocr Metab 1994;78:1308-1312.

Malley J, Hornstein J. Quantum statistical inference. Statistical Sci 1993;8;433-457.

Malley J. Statistical applications of Jordan algebras. Research monograph in *Lecture Notes in Statistics Series,* vol 91. New York: Springer–Verlag, 1994.

Presentations

Malley J. "Quantum Statistical Inference," an invited lecture series, Department of Statistics, Harvard University; April 28-29, 1994.

Malley J. "Quantum Statistical Inference," a two week invited lecture series, Institute of Mathematics/Institute of Theoretical Statistics, Aarhus University, Denmark; August 17-30, 1994.

Computational Methodology Collaborations and Support

James DeLeo

The DSB uses computational methodology to provide systems analysis, design, and programming services to support biomedical research, clinical decision-making, and administrative problem solving. Computational methods currently employed include Receiver Operating Characteristic (ROC) methodology, neural networks, image processing and general PC-based biomedical applications programming. Services are provided through consultations, collaborations, classes and seminars, and special interest groups.

Numerous consultations concerning the application of ROC and neural network methodologies were provided this year. Several active collaborations are in progress including: a pharmacy patient interviewing system with the Clinical Center's Pharmacy Department; development and implementation of ROC methodology for biomedical applications with NINDS (Greg Campbell, DIR); continuing development of ROCLAB software with NINDS and NCI (Greg Campbell and Michael Debaun, DCE); identifying "atrophy" regions in Positron Emission Tomography (PET) scans in support of research in Alzheimer's disease and other forms of dementia for NIA (John Van Meter, LN).

Several classes and seminars were presented on the following topics: ROC methodology; neural network methodology; and introductory programming. DeLeo also provided coordination for an active campus-wide Neural Network Interest Group.

Support for Laboratory Data Acquisition and Analysis

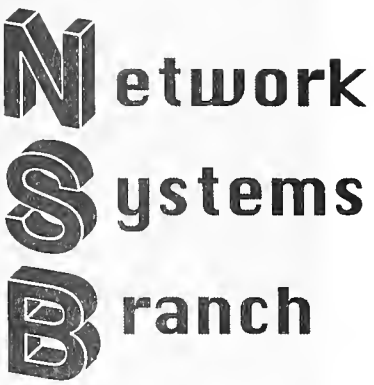
Hal Fredrickson

DSB staff conducted an evaluation with NIDDK of LabNotes, a Microsoft Access-based system developed to track lab data. The NIDDK Hepatitis Study Section evaluation involves studies of various types of hepatitis; this product is being evaluated for general use at NIH.

DSB staff continued support of several PC-based laboratory automation projects. In cooperation with the Biomedical Engineering and Instrumentation Program (BEIP) of the National Center for Research Resources, we provide computer interfaces and software support for NIH developed laboratory instruments such as a 100 Channel High Speed Spectrophotometer. This year the spectrophotometer system was enhanced with a higher resolution timer, higher resolution stop flow, and a higher speed liquid crystal shutter. These enhancements allow the computer to control the laser, start data collection, stop flow (control chemical mix) and control two other variables with more flexibility.

Beginning in 1976, DCRT developed and installed 11 Laboratory Data Acquisition and Control System (LDACS) computer systems throughout Building 2 for NIDDK. Based on DEC LSI-11 microcomputers, the LDACS computers were connected to laboratory instruments for control and data collection. We are now updating the LDACSs with more modern technology: DSB staff have replaced three LDACS computers with various IBM PCs. Requests for more LDACS replacements are being considered.

Several of these projects will be phased into the newly formed Laboratory and Clinical Applications Section.


Network
Systems
Branch

Network Systems Branch

Harold Ostrow, Chief

Network connectivity has become an essential tool for the biomedical, clinical, and administrative communities at NIH. The NIH wide area network, known as NIHnet, serves to interconnect local area networks (LANs) throughout NIH, providing the data highway over which images, data and electronic mail can travel. All NIH's institutes, centers, and divisions have LANs that depend on NIHnet for wide area network services. The Network Systems Branch (NSB) had a number of important accomplishments during FY94, including:

- extension of optical fiber for the high-speed network backbone to buildings throughout the NIH campus, resulting in substantial increases in network capacity and reliability for many network users
- installation of high-speed fiber optic network connections to several off-campus sites, replacing older technology telecommunications links
- completion of the NSB-designed network infrastructures for new Buildings 49 and Natcher, and development of a network wiring plan for the Rockledge Building
- announcement of production support for the Microsoft® Mail hub and gateway, providing full service to MS Mail users, including e-mail to and from the Internet
- addition of nearly 40 NIH LANs to the NIHnet metropolitan area network, bringing network access to NIH and Internet services for hundreds of NIH employees
- introduction of new network router and LAN hub platforms to the NIHnet infrastructure, which incorporate new capabilities into the state-of-the-art network.

NSB Overview

The Network Systems Branch (NSB) has a mandate that includes the following functions:

- design, implement and monitor the high-speed backbone connections that provide connectivity to local area networks (LANs) on campus
- coordinate, implement and monitor NIHnet connections to NIH LANs in off-campus buildings and to other networks
- provide guidance and support for locally managed LANs and for "whole building" wiring infrastructure
- design and support DCRT networks, including networks for trans-NIH services and servers
- develop information resources and network-based applications for the NIH community.

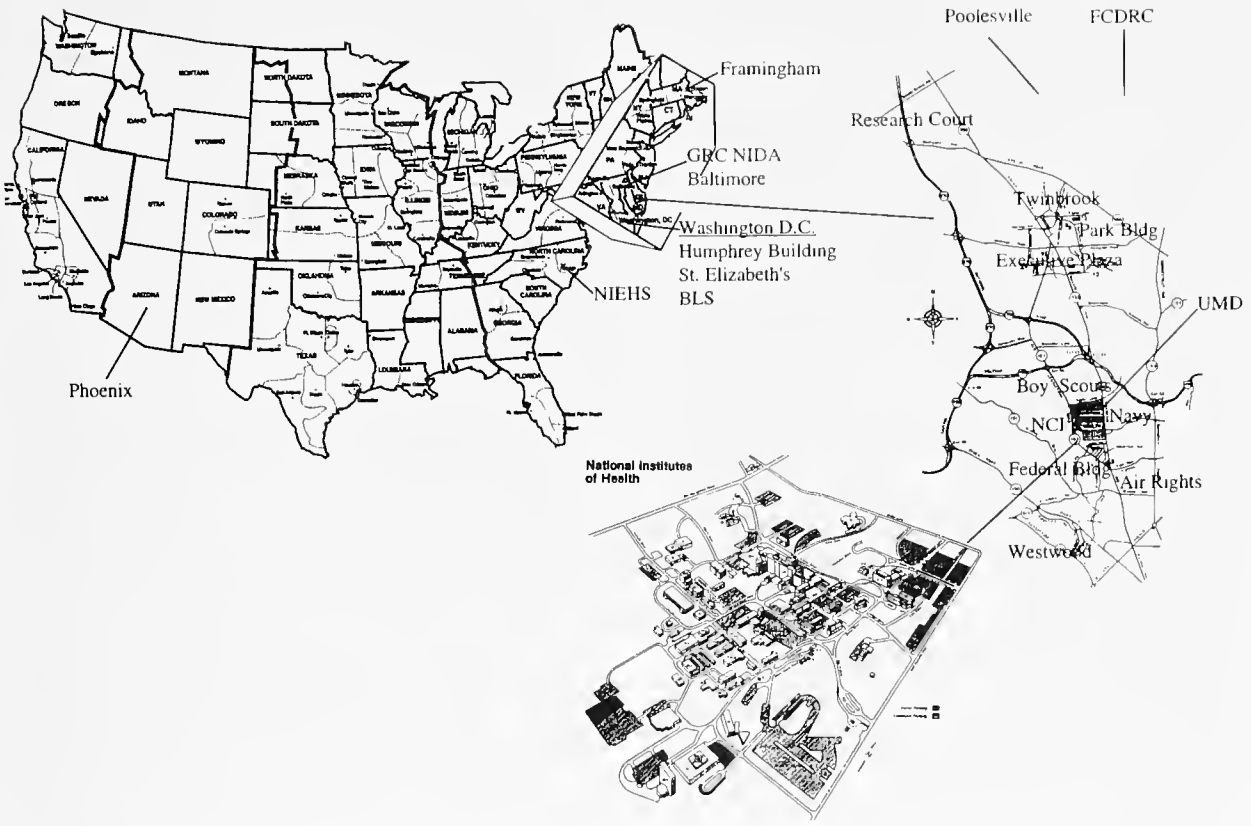
The Network Systems Branch is composed of three sections: the *Network Research and Development Section*, the *Customer Support Section*, and the *Systems Development and Support Section*. The branch's professional staff includes electronics engineers, computer scientists and computer specialists. In addition to the efforts of each section, the entire branch participates in network monitoring and handle questions and problem reports on the NIHnet Customer Support "hot line."

NIHnet Leverages High Speed Fiber and Existing Telecommunications

The NIHnet infrastructure ties together over 220 LANs that exist in the Clinical Center, in other on-campus buildings and in numerous off-campus buildings. This wide range of locations, coupled with the variety inherent in mixed scientific and administrative use, requires flexibility, reliability and responsiveness on the part of NSB and the NIHnet wiring plant itself. Multiple connectivity strategies are integrated into NIHnet: an on-campus high-speed fiber backbone, on-campus Ethernet over fiber connections, wide area fiber connections and lower speed telecommunications-based wide area connections.

The high-speed fiber backbone, which utilizes the 100-megabit-per-second FDDI (Fiber Distributed Data Interface) technology, provides the NIHnet connection for over 150 LANs in on-campus buildings. The FDDI connections to Buildings 1, 6, 9, 14, 21, 31, 35, 38, 38A, and the Children's Inn were installed during FY94. Physically, the fiber optic cabling used in the FDDI backbone is in a "hub and spoke" configuration, with concentrators in controlled environment areas in buildings 10, 31, and 12. Logically, however, FDDI utilizes a dual ring structure, which enhances the robustness, fault tolerance, and redundancy of the backbone. The FDDI backbone has proven to be an extremely reliable, high capacity foundation upon which to build the interconnections between on-campus network intensive locations.

The buildings served by the FDDI backbone house a large percentage of the intramural scientists on campus, whose research spans the spectrum of NIH endeavors, and whose research can have data-intensive networking requirements. Imaging, full-motion video, genome mapping, and other research applications place heavy demands on the network, with data transmissions of a time-sensitive and "bursty" nature. Furthermore, in the next few years, desk-to-desk videoconferencing is



NIHnet connects Local Area Networks at NIH, in the Washington, DC, area, and across the country.

likely to result in explosive growth in demand for network services.

The FDDI backbone also serves as the NIHnet connection for many of the largest administrative and extramural LANs on campus. These LANs often serve several hundred users, and require the kind of reliable production connectivity that the FDDI has demonstrated. Furthermore, the high capacity FDDI backbone provides the bandwidth to handle the upcoming generation of client-server database applications in the grants administration area.

NSB's goal is to provide, via the high-speed FDDI backbone, both the NIH scientific and administrative communities with a reliable network infrastructure that has sufficient capacity to meet their needs within tactical time frames. Future fiber based technologies that provide additional network capacity, such as Asynchronous Transfer Mode (ATM), would use the same fiber infrastructure currently being used for the FDDI connections.

In buildings that have smaller populations of network users, the NIHnet strategy of choice is to connect the LAN to NIHnet in another building using the fiber infrastructure. This type of connection provides the same dedicated, high capacity connection between the LANs in the building and NIHnet's FDDI backbone, while being very cost effective. During FY94, buildings 3, 5, 8, 15K, 16, 32, 41, 60, and the Children's Inn were connected to NIHnet by Ethernet-over-fiber connections, which can be up to 2 kilometers in length. This way, even buildings with one or two LANs can participate in NIHnet with high capacity connections. FY94 also saw the implementation of high speed fiber-based connections to off-campus NIHnet sites. Ten megabit-per-second connections, using Bell Atlantic's Fiber Network Services (FNS), were installed in the following locations:

- Executive Plaza North/South complex, 6120/6130 Executive Plaza Blvd., Rockville, MD
- Gateway Building, 7201 Wisconsin Ave., Bethesda, MD

- Parklawn Building, 5600 Fishers Lane, Rockville, MD
- Federal Building, 7550 Wisconsin Ave., Bethesda, MD
- 6000 Executive Blvd., Rockville, MD
- 6006 Executive Blvd., Rockville, MD
- SURAnet Regional office in College Park, MD.

FNS provides greater reliability and additional capacity for NIHnet off-campus connections at a price that is competitive with existing telecommunications technologies.

One of the most heavily utilized NIHnet connections is with SURAnet's office in College Park, Maryland, the regional component of the Internet. All NIHnet traffic to the Internet, including file transfers and interactive sessions on distant machines, goes through the SURAnet connection. In view of the current high capacity data transfer done on this line and in anticipation of even heavier traffic in the near future, the NSB established an FNS link to SURAnet with 10 megabits per second capacity in FY94. This way, NIHnet's "on-ramp" to the "Information Highway" will be able to meet capacity requirements for the near future.

NIHnet utilizes existing telecommunications technologies and infrastructure to provide flexibility and meet off-campus networking requirements. Over 70 LANs are connected to NIHnet, using 1.5 megabit per second ("T1") telecommunications lines. T1 communications can usually be transmitted over existing telephone lines, which eliminates the need for additional wiring and makes possible timely and cost-effective installation of new NIHnet connections. Temporary locations on campus, including construction trailers used by the Division of Engineering Services, are served by T1 connections. In addition, T1 lines are well suited for off-campus connections, such as the connections to the National Institute of Environmental Health Sciences in Research Triangle Park, NC, NIH locations on Executive Boulevard, and some NIH contractor sites. In some cases T1 connections are available as a subset of the D3 telecommunications capacity that goes to off-campus NIH buildings (e.g., 6100 Executive Blvd., the Animal Care complex in Poolesville, and the Mail/Print facility in Rockville).

NIHnet Monitoring and Support

The Network Systems Branch extensively monitors NIHnet to ensure continuous and reliable service for users on connected LANs. NSB staff have network monitoring stations, software LAN analyzers and hardware diagnostic tools to assist in this process. LAN Coordinators are contacted in the event of network problems to assist in the diagnosis and solution of the problems.

The network monitoring stations used by NSB display a map of NIHnet, showing correctly functioning connections in green. If a connection to a LAN stops communicating to NIHnet, the network monitor sounds an alarm and changes the color of the connection on the network map to red. This way, NSB staff are immediately alerted to problems, and can contact the LAN coordinator and start the "problem diagnosis and resolution" process quickly. An additional benefit of the NSB's proactive stance is that the NSB staffer's call often gives the LAN coordinator a "heads up" warning that a local LAN problem may be occurring. The NSB is working to further enhance and expand these network monitoring capabilities and to keep our installation at the state of the art.

NSB provides mechanisms for LAN coordinators to initiate NIHnet questions or problem reports. Perhaps the most popular is the NIHnet Customer Support "hot line," a telephone consulting service that is available from 8:30 a.m. to 4:30 p.m. each work day and is staffed by senior networking professionals. On a typical day, the hot line receives between 10 and 15 calls on topics such as LAN wiring recommendations, Grateful Med installation, IP (Internet Protocol) address registration in the Domain Name Server, electronic mail addressing, how to connect to a UNIX® server, and how to determine if a PC is correctly communicating over the network. The NSB is working with DCRT's Customer Services Branch (CSB) to bring the NIHnet hot line under the umbrella of the "4-DCRT" Central Point of Contact service; this service enhancement will take place early in FY95.

The NSB also encourages problem reports or questions via electronic mail, to the NIHNET@LIST.NIH.GOV address. This way, a LAN coordinator can send complete problem documentation directly to the NSB for assignment to the most appropriate networking expert.

As a way to establish and maintain close contact with the LAN coordinators as a group, the NSB participates in the Campus User Research Exchange (CURE) meeting. The CURE meeting provides a forum for the NSB and other DCRT groups to make direct contact with networking leaders and Technical LAN Coordinators (TLCs) throughout NIH for an open and public exchange of ideas, comments, questions and concerns. In FY94, the NSB made a number of presentations at the CURE meeting, including

Microsoft® Mail, IPX routing and the NIHnet topology.

The NSB is consolidating the inventory of NIHnet components, contacts and events into a comprehensive database based on the Action Remote® system from Remedy Corp. A single server hosts a problem and incident tracking system, a database of routers, lines and LAN connections, and a database of LAN coordinators and other network contacts. This consolidation facilitates effective tracking of incoming calls to the NIHnet hot line and allows for statistical analysis of network performance, capacity, and reliability. Current plans are to integrate the NSB's Remedy system with the Customer Services Branch's system based on the same technology, as a way to unify problem tracking throughout DCRT.

Preparation for the Unexpected

Providing a networking infrastructure for NIH that is reliable and sustainable requires preparation for unscheduled network events. The NSB regularly deploys Uninterruptible Power Supplies (UPSs) for hub and outboard routers, so that transient power spikes or drops will not knock out or damage NIHnet routing equipment.

Other steps are taken to prevent, minimize and diagnose network outages: the FDDI backbone is backed up by an Ethernet in the event of a ring failure; backup T1 interfaces are configured on the hub routers for quick recovery from hardware failures; and routers are equipped for remote diagnosis via telephone connections in the event of network problems. All these precautionary features were called upon during FY94 to respond to a host of incidents or events involving NIHnet.

In FY94, the NSB deployed a new generation of routers, the Cisco 7000 series, in the Clinical Center; these provide "hot swappable" components, which will further reduce any downtime due to hardware problems. Additional deployments of the Cisco 7000 family of routers at large NIHnet sites are planned.

Network Guidance and "Whole Building" Wiring

From the LAN perspective, management is easier and more effective, and reliability is higher when the LAN wiring is well designed and properly installed. From the wide area network perspective, NIHnet is more robust and easier to support if connected LANs are well managed. Accordingly, the NSB provides extensive consulting to NIH groups taking the first step into networking. In a team approach with the Customer Services Branch (CSB) and Distributed Systems Branch (DSB), NSB staff provide "organizational consults" to help define network strategies, technologies, and applications to groups that are preparing for networking.

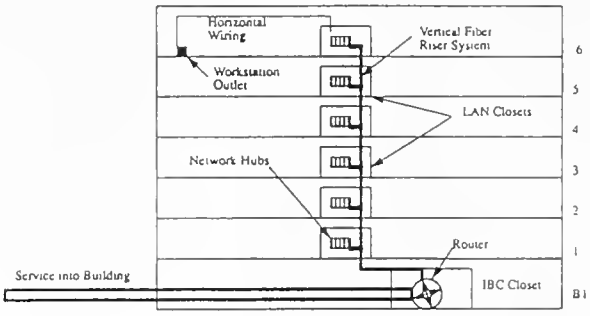
Organizational consults and advice on network wiring strategies are provided for large groups as well as small groups. For example, in FY94 the NSB consulted on networking issues related to the upcoming move of NHLBI, DRG, OD and other personnel into the Rockledge complex in Rockville. By developing a building-wide strategy, there are standards of installation and economies of scale for all the groups moving into those buildings. The NSB's input on standardized network wiring in the new Natcher and rental Rockledge Buildings is now coming to fruition, as the buildings near completion and occupancy. By providing networking assistance during the design phase, NSB encourages standardization in the network wiring plant at NIH and, as a result, helps ICDs avoid expensive retrofitting and rewiring.

As each of the ICDs move into the buildings, they find they can simply "plug in" their computers and obtain the entire gamut of networking services. This would not have been possible without the behind-the-scenes activities of the NSB, fully coordinated with the Telecommunications Branch of ORS and the Division of Engineering Services. The Rockledge and Natcher buildings contain "vertical" and "horizontal" fiber optic cable, permitting up to 100-fold increases in capacity, when needed, without the need for rewiring. NSB contributed in an essential way to everything from the the placement of cable trays, to protect the investment in fiber and cable, to monitoring the quality of the installation and quality control.

DCRT's Networks

The Network Systems Branch is responsible for maintaining LANs for DCRT in many buildings. NSB has designed and maintains Ethernet-based networks in the Building 12 complex, the Building 12 Machine room, Building 31 and 6100 Executive Blvd. The Building 31 network uses Thin Coaxial cable, the LAN in 6100 Executive Blvd. uses unshielded twisted pair, while the Machine room Ethernet and the main DCRT Ethernet in Building 12A use a combination of Thick Coaxial Cable and Unshielded Twisted Pair. The Thick cable is difficult to manage and has reached the end of its life span. The NSB has developed a new wiring plan that will support the current and future high speed data communications needs for DCRT and the rest of the NIH campus.

Over the past year the DCRT Ethernet has experienced a rapid increase in the number of nodes attached (over 500) and now utilizes the complete available bandwidth. The implementation of the new network wiring plan was delayed until asbestos ceiling



DCRT designs "whole building" wiring plans for new and refurbished buildings, with specifications for LAN Closets, vertical optical fiber connecting the Closets, and horizontal wiring to individual offices.

tiles could be removed. As a stopgap, short-term solution, the DCRT Ethernet was split into two separate networks. In FY95, a new cable plant will be installed in the 12A complex. This new cable plant will provide flexible expansion, central management capabilities, easier network troubleshooting, and connectivity among a wide variety of systems and networks (e.g., Ethernet, Token Ring, FDDI), as well as an infrastructure for upgrading phone service to the complex.

The rewiring plan calls for a star wired cable plant, consisting of unshielded twisted pair and fiber optic cables to meet the requirements of current and future high speed LAN technologies and standards. A central wiring closet will house network hubs, servers and patch panels, and will serve as a centralized network control center. Prewired data outlets in all offices and work areas will allow for easy network access, reconfiguration and expansion.

During FY94, the Machine Room Ethernet, which connects production service computers, such as the IBM mainframe and e-mail hubs to NIHnet, was upgraded to improve capacity, reliability, and monitoring capabilities. Unshielded twisted pair and high capacity hubs replaced thick coaxial cabling in the upgrade; separate services were also provided via dedicated Ethernets, so that traffic from the Microsoft® Mail hub/gateway is now largely segregated from traffic from the IBM mainframe services. The flexibility inherent in the new Machine Room network topology will allow for future capacity upgrades to meet the network requirements of centralized production servers, such as database gateways, domain name servers, and information servers.

NSB Provides NIHnet-Wide Network Services

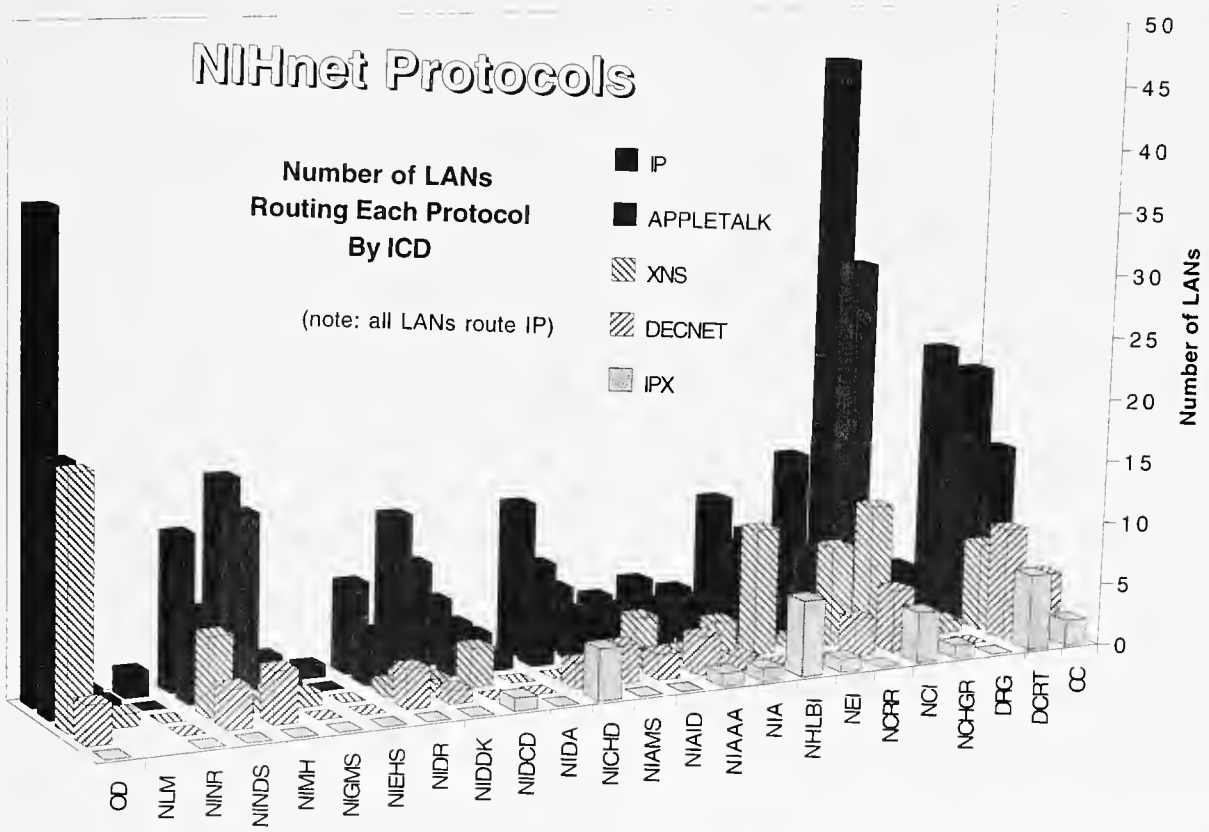
In addition to planning, deploying, and supporting the NIHnet infrastructure, the NSB also develops information resources and network-based applications. These network services add value to NIHnet for the NIH community.

Electronic mail gateways provide one of the most essential network services, because the gateways allow users to correspond across the network with colleagues and collaborators. NSB supports two widely used electronic mail gateways: the Microsoft® Mail gateway; and the 3+Mail gateway.

DCRT recommends Microsoft® Mail for larger or centrally supported LANs, and supplies MS Mail server software to ICD LAN administrators. NSB runs the Microsoft® Mail hub and gateway, which moves mail between MS Mail LANs, and converts mail from an individual LAN into the Internet mail standard format, known as the Simple Mail Transport Protocol (SMTP), for transmission to other LANs, NIH UNIX® or mainframe users, or remote Internet sites. Currently, Microsoft® LAN Manager LANs, Microsoft® NTAS LANs, and Novell Netware LANs are using Microsoft® Mail and are connected to the Microsoft® Mail gateway.

The Microsoft® Mail gateway, which received full production status in January 1994, currently handles mail from 80 LANs at NIH. A number of enhancements were implemented on the Microsoft® Mail gateway during FY94. Synchronization of user e-mail address directories between Microsoft® Mail servers, along with user address exchange with the NIH e-mail directory, has resulted in complete NIH-wide e-mail directory information being available to e-mail users. In addition, "bullet-proof" backup and recovery systems and additional operational monitoring were put into production service. The NSB is working closely with other DCRT groups, ICD LANs, and the Microsoft® Corporation to ensure that the Microsoft® Mail gateway is an effective, long-term electronic mail distribution mechanism for cross-platform and trans-organizational communication.

DCRT is phasing out the 3+Mail gateway, which supports legacy 3Com 3+ LANs at NIH. As part of this phasing out, DCRT is actively assisting ICD LAN



Number of network protocols routed via NIHnet to NIH Institute, Center and Division Local Area Networks

administrators in their move to other e-mail systems. While it is clear that 3+Mail usage at NIH has dwindled considerably, the NSB will continue to support the 3+Mail gateway in early FY95 to meet ICD LAN requirements, but will discontinue the service later in FY95.

Several other e-mail-related projects are under way. NSB is working on providing an NIHnet-wide fax gateway to replace the 3+Mail-based fax gateway currently on PUBnet. A fax gateway accepts e-mail sent from an NIHnet-connected workstation, and converts it to a fax to be sent over conventional telephone lines to the recipient's fax machine. This service makes it possible, for example, to include fax destinations in the CC (Courtesy Copy) list for electronic mail, sending the e-mail via fax machine to people who are not network-connected.

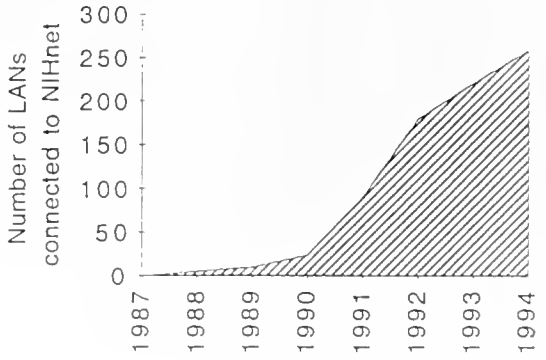
NSB is testing the IMAP2 protocol, which is a client-server protocol for Internet mail. It is similar to the POP3 protocol currently used by DCRT's "popserver" service, but provides additional features. This e-mail system testing is an element of NSB's participation in DCRT's E-mail Directions group.

Dial-up access to the network will be an area of interest for NSB during FY95. NSB is working to provide dial-up network access to the NIH community. This project will allow NIHnet users at home and while traveling to obtain access to NIHnet via a dial-up connection from a PC or Macintosh.® Connections via the PPP (Point-to-Point Protocol for TCP/IP) will be supported; PPP is the preferred standard for effective dial-up network access. The goal of the NSB is to work with other DCRT groups to provide network access to NIHnet users who happen to be away from their regular NIHnet-connected workstation. Dial-up access is not intended to replace workstation connections to NIHnet, but rather to augment those connections and to provide more flexibility to the NIHnet access mechanisms available to the user community. A major aspect of the development effort is handling of authentication of users and other security issues.

NSB collaborates with other components of DCRT

to help provide other network services:

- with OD on GOPHER, MOSAIC, and WorldWideWeb services
- with the Computer Facilities Branch (CFB) on the development of a comprehensive NIH e-mail directory
- with the Distributed Systems Branch (DSB) and CFB on the Microsoft® Mail-to-NIH e-mail directory information swapping
- with CSB on a network-based problem tracking system for use by the Division and the NIH community.



Growth in the number of NIH Local Area Networks connected to NIHnet from 1987 to 1994.

In FY94, NSB announced support for the Ushare software system that allows TCP/IP LPR printer clients to print on AppleTalk printers. This is useful for UNIX® systems, PC LANs not using a network operating system (NOS), and IBM mainframe printing (including Delpro), with the Interlink EPS product being tested by CFB. Ushare operates on a Sparcserver-10 UNIX® systems, which resides in the Machine Room. NSB is investigating the possibility of providing a similar service for Novell® Netware printers.

Technology Tracking for Today and Tomorrow

Computer networking is a volatile industry, with new networking companies and technologies and even philosophies appearing almost daily. The race between new products with higher capacities and new applications with higher capacity requirements is a perpetual dead heat. In view of this, the NSB actively tracks and influences networking trends, technologies, and standards, with the goal of maintaining and augmenting the capacity and sustainability of NIHnet. This must be done within the framework of current technologies, while preparing for deployment of upcoming technologies, and maintaining cost-effectiveness for the entire community.

A good example is NSB's recommendation that NIHnet LANs cable for Ethernet in a star and hub "10baseT" configuration using Type 5 wire, which provides 10 megabits per second (10 mbs) capacity. This conforms to the current industry standard, with benefits including very low cost for Ethernet cards and better diagnostic maintenance. In addition, this recommended configuration positions NIHnet LANs for the upcoming 100-megabits-per-second (mbs) Ethernet technologies that are now starting to be deployed. This will allow ICD LANs to leverage their current investment when implementing the next generation technology. NSB is closely tracking the 100 mbs Ethernet technology as a mechanism for handling high bandwidth scientific applications, such as medical imaging, full-motion video, data visualization, and 3D computer modeling. NSB anticipates putting some servers on 100 mbs Ethernet during FY95.

The NSB employs state-of-the-art FDDI technology on NIHnet's campus fiber backbone. The FDDI standard has been ratified by international standards bodies, and is implemented by a number of networking equipment vendors. As a result, market forces decrease the cost of FDDI implementation, to the benefit of NIHnet users. As an additional benefit, the fiber infrastructure in place for FDDI will also be compatible with the next generation of backbone technology: Asynchronous Transfer Mode (ATM). During FY95, the NSB will prototype an ATM network, in anticipation of the commercial availability of standards-based ATM hardware and software and of expected increases in demand for capacity in the NIHnet backbone. By deploying a standards-based FDDI backbone now and tracking the ATM technology, the NSB is meeting today's needs while positioning the NIHnet backbone for tomorrow's technologies.

In FY95, the NSB will also track network-based applications in order to forecast future NIHnet connection and capacity requirements in upcoming years. In areas as disparate as molecular biology and personnel management, more and more applications rely on the network for communications. The network-based applications provide the major impetus for the NIH scientific and administrative communities to become connected to NIHnet. For example, as client-server database applications are implemented, such as DRG's IMPAC system replacement with its related CRISP and EGAD components, it is anticipated that

large new audiences will request NIHnet connections. Other network-based services are also expected to generate new demand for additional NIHnet connections and capacity:

- Information services such as GOPHER, MOSAIC, and Grateful Med
- Medical imaging in clinical research
- Laboratory imaging
- Access to supercomputers
- Data visualization
- Clusters of high performance workstations.

NSB fosters substantial personal contacts with the various communities throughout NIH, to "keep a finger on the pulse" of the network requirements for upcoming applications. In addition, NSB will continue to deal with high priority issues:

- rapid growth
- heterogeneity of systems, users, applications, and requirements
- rapidly changing technology
- requirements for monitoring, redundancy, and capacity management
- aging buildings, some containing hazardous materials, in which a network presence is required
- delays in procurements and in construction projects
- personnel constraints in the area of LAN and NIHnet support
- need for collaboration with NIH organizations such as OIRM, TCB, PHS Telecommunications, DES, and ICD LAN support groups.

The NIHnet infrastructure will continue to grow as the requirements for network connectivity grow at NIH. The NSB recognizes that crucial production applications in the clinical, intramural, extramural, and administrative areas depend on a robust, reliable network. Through commitment to quality, strategic planning, customer support, and proven technology, NSB continues to provide state of the art network service to the NIH community.

Information

Systems

Branch

Information Systems Branch

The Information Systems Branch (ISB) continues to be a central NIH resource that provides advice and services to the NIH user community in the development and maintenance of computer based information systems. The ISB provides advice and assistance to researchers, program officials and administrators throughout NIH in planning for and obtaining computer information services. The branch also develops, maintains, and processes the NIH Administrative Data Base and the Clinical Center's Clinical Information Utility. On the staff are 37 permanent full-time employees, whose disciplines include computer science, mathematics, and information systems.

The branch is composed of four sections:

- The *Applied Systems Programming Section* (ASPS), led by Larry Martin, provides general analysis, design and programming services to the NIH community. This includes custom application development services for various platforms and configurations.
- The *New Technology Analysis Section* (NTAS) tracks and evaluates rapidly evolving database technologies and provides advice and guidance for implementation of new technologies within the NIH computing infrastructure. This section is responsible for analyzing and selecting new database management approaches and for developing the techniques that will facilitate their use across multiple platforms. Focus is on distributed client/server computing, including frontware technologies, query and reporting tools, application development tools, database servers, and database connectivity software.
- The *Data Base Applications Section* (DBAS), led by Barry Madia, is responsible for the development and maintenance of the Administrative Data Base System, which provides broad support for all administrative and financial processes at the NIH.
- The *Data Base Information Section* (DBIS), led by John Price, is responsible for providing the NIH user community with information stored in the Administrative Data Base System. This takes the form of batch reporting as well as online *adhoc* queries using graphical user interface, client/server, and relational data base technologies.

The NIH Administrative Data Base Supports the NIH Mission

The NIH Administrative Data Base (ADB) represents a major effort by the NIH to combine the administrative and financial data of its intramural program. Using an integrated approach to data base management, the ADB concentrates on the full sharing of data among all subsystems that support the NIH intramural program. It features online point-of-origin data entry, minimized data redundancy, background generation of all accounting transactions, and fully synchronized information processing, i.e., the user will always obtain the latest state of any process, data, or function.

The development of the ADB is an ongoing project that encompasses the following features:

- the purchasing, receiving, and payment of goods and services is fully supported
- items in nine individually tracked inventories are made available by way of online stock requisitions and are completely integrated with the operation of the self-service stores
- all vendors, vendor credits and vendor source agreements are maintained and tracked
- NIH cashier functions are fully supported
- eleven service and supply fund activities have been integrated
- foreign, domestic, local and Clinical Center patient travel orders and travel vouchers are processed and tracked
- the AIDS Loan Repayment System, to support the repayment of outstanding student loans for scientists who are conducting AIDS-related research at NIH, is integrated into the ADB and utilizes its procurement, invoice and accounts payable functions
- an NIH-wide property management system is functional with data capture initiated by the receiving module
- the implementation of full research contracts support is under development, and accounting functions such as fund formulation and funds certification have been shifted to online ADB support.

A more specific summary of new ADB initiatives during FY94 is presented below:

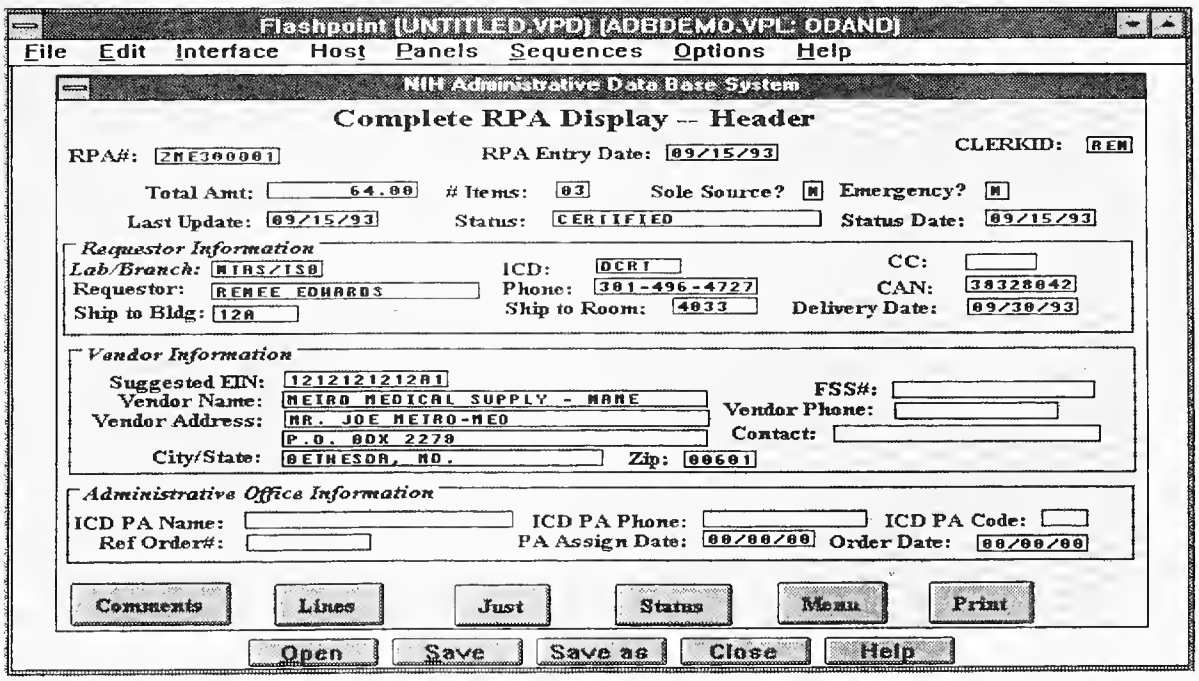
- *Service and Supply Fund Activity System* (SSFAS). Phase I of the Division of Engineering Services (DES) activity implemented work request tracking, funds approval, status display, and reporting. A requirements analysis and design document to support Phase II for this activity were developed in FY94. Phase II will add all Central Accounting System (CAS) functions (obligations and disbursements), including Common Accounting Number (CAN) changes, DES labor

billing, inventory billing, special purchases billing and reporting functions. An evaluation of options for accepting Government Printing Office (GPO) invoices in three digitized formats rather than paper form was conducted for the Printing and Reproduction Branch (PRB) during FY94. These new formats consist of magnetic tape, PC diskette, and an electronic Bulletin Board. This capability eliminates the need for manual data entry into the ADB for processing of GPO invoices by the PRB staff. A requirements analysis was performed to support Vetinary Resources Program (VAP) activity within the NIH ADB. The system design and development effort will begin as resources become available.

- *Travel Management.* The computer software to support Sponsored Travel was developed and tested during this fiscal year. A group of ADB users was established to participate in a pilot test for the system. Upon conclusion of the pilot test, this function will be available to the entire NIH community.
- *Financial Management.* An electronic transmission of NIH payment data to the

3270 Screen vs. Graphical Interface

```
                        COMPLETE RPA DISPLAY --  HEADER                    ODAND

RPA #: ZME300001              RPA ENTRY DATE: 09/15/93              CLERKID: REN
LAB/BRANCH: NTAS/ISB             ICD: DCRT              CC:         USERID: REN
REQUESTOR: RENEE EDWARDS         PHONE: 301-496-4727      CAN: 38328042
SHIP TO BLDG: 12A               SHIP TO ROOM: 4033      DELIVERY DATE: 09/30/93
SUGGESTED EIN: 1212121212A1     TOT $ AMT:        64.00      # LINES: 01
VENDOR NAME: METRO MEDICAL SUPPLY - NAME            FSS#:
VENDOR ADDRESS: MR. JOE METRO-MED                   VEN PHONE:
                P.O. BOX 2278                    CONTACT:
        CITY/ST: BETHESDA, MD.             ZIP: 00601   SOLE SRCE?: N EMRGNCY?: N

LAST UPDATE: 09/15/93     STATUS: CERTIFIED              STATUS DATE: 09/15/93

***                     ADMINISTRATIVE OFFICE SECTION                       ***

ICD PA NAME:                      ICD PA PHONE:              ICD PA CODE:
        REF ORDER#:            PA ASSIGN DATE:  00/00/00 ORDER DATE:  00/00/00

COMMENTS:


PF3: LINES  PF4: JUSTIFICATION  PF5: TO STATUS  PF7: MENU  PF9: PRINT
```



Department of Treasury located in Philadelphia, PA, was implemented in collaboration with staff from OGFM, DHHS and the U.S. Treasury.

- *Loan Repayment Systems.* During FY94, a requirements analysis was conducted to incorporate additional types of loan repayments such as Clinical Research, Contraception/Infertility, General Research, and the Undergraduate Scholarship Program. Implementation of these types of loan repayments is scheduled for FY95.
- *Radioactive Material Ordering System* (RMOS). The Radioactive Material Ordering System will provide consolidation of ICD radioactive material orders so that selected vendors receive one order per day from NIH, volume discounts to ICD users as order quantities increase during the day, online catalog item selection and pricing, generation of accounting transactions, collection of NIH-88 information for Radiation Safety Branch (RSB) tracking, centralization of receiving of the consolidated orders, ICD delivery information for ordered materials. RBS will pilot test this system in early January, 1995.
- *Fellowship Payment System* (FPS). During FY94, the requirements analysis, system design and computer programs were completed to support the Fellowship Payroll System Phase I as part of the ADB. The FPS will support registration of Fellows by Fogarty International Center staff, processing renewals, extensions, absence without stipend, reactivation of fellows, early terminations, normal terminations and ICD transfers, printing of PHS-1485 (Approval List), PHS-4165 (Fellowship Activation Notice), release of obligation transactions to NIH Central Accounting System (CAS), downloading of data to DFM for further processing of payroll actions and payments.
- *Property Management.* New procedures for reconciling the annual property inventory were developed during this fiscal year. A requirements analysis was also conducted to interface the Property Management Information System (PMIS) with CAS.
- *Chief Financial Officer* (CFO) *Audit.* In response to the FY93 CFO auditor's suggestion that ADB security be improved, several enhancements were made. The ADB now permits multiple approval officials within a node. Previously, the approval was limited to a single individual in each node. Using the new feature, the primary approval official may designate backup approval official(s) who may be authorized to approve requests in all, or a limited subset of ADB sub-systems: Delegated Procurement; Market Requisition; Stock Requisition; Service and Supply Fund Activity; Work Requests; Travel Order Advances and Vouchers. An electronic registration/deregistration function for ADB access was developed and tested. Using this function, the approving official will be able to review his/her user list and its associated security profiles; request new user identification codes; request changes to existing user security profiles; and deactivate a user.

Administrative Data Base Information System (ADBIS) Provides Easy Access to ADB Information

In collaboration with over 70 ICD representatives, the ISB staff developed the ADB Information System (ADBIS) to provide timely and accurate information from the ADB and CAS to the NIH user community. The ADBIS currently supports Obligations/ Commitments, Procurements, Market Requisitions, Service and Supply Fund Activities, Stock Requisitioning and Central Stores, and Travel Summary information. The Travel Detail and Property functions are currently being developed and should be available during the fourth quarter of FY94. Contract staff have begun to develop a graphical interface for the on-line query and decision support.

Client/Server Development Tools

ISB has continued its efforts to investigate the viability of various client/server tools for the NIH computing infrastructure. As a bridge to client/server computing, ISB evaluated several "frontware" technologies. These products are designed to extend the life of legacy systems by building graphical front-ends to existing applications. Included in this investigation was the Flashpoint product, which is a Windows®-based development tool for creating graphical interfaces and for integrating information at the desktop. This investigation involved developing a graphical user interface (GUI) for a component of the ADB and the ADBIS. Each of these prototypes successfully demonstrated how frontware technologies can be used to complement existing legacy systems and thus provide a more intuitive user interface.

The ADBIS prototype also included the integration of the Graphical Query Language (GQL) product, which is a query and reporting tool that provides cross-platform support for both the Windows® and Macintosh® environments. This component demonstrated the functionality of *ad hoc* data retrieval using a point-and-click interface, an Executive Information System (EIS), and data exchange with

microcomputer-based software packages such as Excel® and WordPerfect® within the ADBIS.

Other ISB-Supported Computer Applications Highlights

Clinical Information Utility

Developed during the 1970s as a historical archive of clinical information for research, the Clinical Information Utility (CIU) gathers data from the Medical Information System (MIS), the Medical Records Department (MRD), and the various service organizations in the Clinical Center (CC). Over the years, millions of records have been archived and made available for use in ongoing research protocols, and for retrospective search and display. The CC Information Systems Department monitors and authorizes all users of CIU data, and the CIU automatically tracks and reports each access of the database. To satisfy clinical investigator needs, the CIU currently handles approximately 10 recurring and 20 *ad hoc* requests each week.

The CIU continues to assist the MRD with a series of studies to track the amount of time patients spend in the hospital for various diseases. In support of ongoing studies in women's health research, the CIU staff assisted the MRD in producing reports that analyze the duration of hospital stay by gender for each institute.

The CIU continues to work with the Medical Record Committee to establish procedures for the presentation of historical laboratory results in System International Units (SIU) in lieu of the current standard lab values, to improve accessibility to any researcher requesting these data. The reports contain information on current units, conversion factor, reference intervals and the calculated SIU values.

In collaboration with the MRD, the CIU staff developed a method to determine the accuracy of data entered by the MRD department. Procedures were developed to compare Discharge Diagnoses data with data from a Discharge Analysis Register to determine whether data have been entered incorrectly or not entered at all.

In collaboration with the Clinical Center and the National Cancer Institute's Clinical Oncology Branch, the CIU staff is investigating options to improve the timeliness and ease of retrieval of data contained in the CIU.

System Modeling

During FY94, a team of ISB systems analysts continued to investigate the use of information modeling as a more structured and less technology-centered approach to systems application development. The mission of the system modeling team was to become educated in the methodologies, techniques, and tools supporting model-driven applications development. Using data, process and logic modeling techniques, the team performed the analysis for a pilot project, Request for Purchase Action (RPA). RPA, a current subsystem of the Administrative Data Base, provides ICD staff the capability of electronically submitting requests for goods and services to their respective ICD ordering office. The Bachman Analyst CASE tool was used by the team to implement the modeling techniques and automate the graphical diagramming processes. The major goals of the modeling effort were:

- to establish a well-documented methodology, which could perhaps serve as a standard for future development efforts within the branch
- to generate documentation that would be appropriate for confirming project requirements with an end user
- to produce a programming specification document that could be passed to a programmer to be manually implemented in the target language of choice.

The System Modeling project has been completed, and a document detailing the experiences and findings of the system modeling team is being finalized. Presentation slides and handout materials were distributed at a formal seminar given in September 1994.

Pilot Client/Server Application for the Clinical Center Medical Records Department

The New Technology Analysis Section (NTAS) successfully implemented two LAN-based client/server solutions to meet the strategic and business goals of the Clinical Center's Medical Record Department (MRD). The collaborative project was part of the NTAS's evaluation of client/server architectural strategies that integrate multivendor software into a total system. Several client/server strategies were prototyped. When the combination of dBASE IV client application development software and the Microsoft® SQL Server database server proved a viable prototype, ISB conducted

a comprehensive requirements analysis and database design effort and produced a pilot application of the client/server technologies in the NIH environment.

Two comprehensive, PC-based information systems were developed to support the daily efforts of the Medical Record Department. The Computerized Microfilm Index System (CMI) is an advanced information system that establishes a relational database and performs the functions necessary to track the process of microfilming inactive and multivolume patient records as well as the reactivation and flowback of medical records. The system provides a single point of entry for the many specialists involved in the tracking process, and coordinates their efforts into a smooth flowing process. To further evaluate client/server technology and support the MRD, a second system, the Medical Record Chargeout System, was developed to automate the record charge-out process of retired records. Although the CMI and the Medical Record Chargeout System are two separate systems, they are merged and presented to the user as a single, total system using relational database and client/server technologies.

The above systems demonstrate the feasibility of using a multi-vendor LAN-based solution to develop systems that provide a GUI front-end to a SQL database. The system has been fully implemented. The start-up phase of implementation included loading data that had accumulated in various formats over the years (1953 to present). The database currently consists of over 180,000 records supported by over 175 programmed procedures and functions. Although the system has been operational for only a few months, the various organizational entities of the department have already noticed improved efficiency in their operations as a result of the common, shared SQL database and supporting technologies.

As part of their client/server paradigm, MRD managers require quick and easy access to their data from both PC and Macintosh® workstations. To this end, ISB has developed a prototype Executive Information System (EIS) for the MRD pilot systems using Andyne's Graphical Query Language (GQL) software. GQL is being evaluated as a client tool for *ad hoc* data querying and reporting. It is a cross-platform tool that will give MRD managers point-and-click access to their data from both their PC and Macintosh workstations.

Collaborative Effort with NIH/OD/Executive Secretariat

ISB's collaborative effort with the NIH/OD/Executive Secretariat was highlighted this year by conversion to newer database application technologies for the Correspondence Tracking System and the introduction of optical scanning and digital image storage and retrieval technology.

The NIH Executive Secretariat (ES) is the receiving, screening, and tracking point for all documents (Executive Correspondence) flowing to and from the Director, NIH: from the Office of the Assistant Secretary for Health (OASH); the Office of the Secretary (OS), other government agencies, the general public, academia, Congress, congressional offices, the White House, and foreign heads of state and their officials. Documents are assigned for action by staff in the ES or disseminated for information. The staff performs editorial and substantive review and rewrite of documents prepared for signature of the Director, NIH, or other high level officials in PHS or HHS, and in other ways contributes to information management at NIH. The Correspondence Tracking System provides support for the tracking and management of correspondence and replies.

During FY94, the system was re-engineered to include many improvements made possible by newer database and networking technologies. The application software was upgraded from interpretive dBASE III to compiled dBASE IV. Improvements were made in the system's ease of use, learnability, maintainability, backup and recovery, and robustness in the networked environment, along with some functional improvements. The new tracking system has been installed for the OD/Executive Secretariat and is currently undergoing testing. Final implementation awaits hardware upgrade acquisitions.

A high point of the year was the introduction of the Executive Secretariat's new optical scanning and digital image storage and retrieval system. The system was developed by Minolta Inc. The ISB provided consultation and guidance to the technical staff of the contractor concerning design and technical issues, and provided for its integration with the Correspondence Tracking System. The digital image retrieval logic depends on data obtained directly from the Correspondence Tracking System. The tracking system was enhanced to make the integration transparent to the OD. Integration of the two systems on the Executive Secretariat LAN was done in collaboration with the staff of NIH/OD/OA/OD LAN.

The modernized Correspondence Tracking System and its integration with the computerized, optical disk correspondence filing and retrieval system provides the NIH Executive Secretariat with a means to provide integrated information management for the NIH. Enhancements under discussion for the coming year

include provision of a more secure and efficient database through client/server technology, and automated, "intelligent" assistance with data entry for specific classes of correspondence.

Child Health Information Portfolio System

The Child Health Information Portfolio System (CHIPS) provides a central facility for timely and easy access to IMPAC and NICHD-specific data for grants, pending applications, jointly funded awards, and subproject and intramural grants for current and all past fiscal years. CHIPS assists NICHD staff by providing tools for the analysis and management of their research grants data.

During FY94 several procedures were developed to eliminate manual processes, improve data integrity, and provide Child Health with more accurate and comprehensive reports for council preparation, study section assignment, and financial commitments. Reporting mechanisms were added to accommodate the newly approved Small Business Technology Transfer grants. A procedure was developed to automatically transmit the dual program assignments captured and maintained by the CHIPS system to DRG for incorporation into the IMPAC system. As a result, program directors will be able to obtain more complete, automated listings of their grant portfolios. Other procedures were developed to automatically log and track program assignments so discrepancies between Child Health and DRG database can be readily identified and their point of origin easily traced. Errors can now be reported to DRG promptly for more timely resolution.

Normal Volunteer Program Tracking System

The Normal Volunteer Program (NVP), sponsored by the Social Work Department, Clinical Center, NIH, provides a central recruiting, entry and tracking function for all paid participants in a variety of NIH clinical protocols. The ISB staff is assisting in the automation of the current manual system. The ISB staff conducted a requirements analysis, reviewed various hardware and software configurations, and established automation objectives. Currently, a PC-based information system is being developed to assist the NVP staff with their information processing needs.

As part of this effort, ISB will design and develop a database management system to collect, query, store, and maintain all information relating to clinical protocols and participants in the NVP at the Clinical Center. The databases will contain: information identifying active and inactive clinical protocols and corresponding medical procedures; protocol recruitment data; census and registration information for each protocol participant; inpatient and outpatient procedure data relating to identified protocols; and compensation data derived from each visit. The system is in the first phase of development. Future plans include possible interface with the CC MIS for additional reporting capabilities and the Central Accounting System for automatic participant payment.

Other Projects

ISB is currently serving as a technical consultant to the Office of Equal Opportunity, offering advice and assistance in the development of the Application Tracking System. This system will be used by NIH EEO offices to evaluate the success of recruitment efforts, the diversity of applicant pools, and the impact of current procedures. Future plans indicate continued DCRT technical involvement.

The ISB continues to collaborate with the NCI Laboratory of Pathology (LP) and the CC to maintain and enhance the NIH Pathology Language Encoding System and the NIH Pathology Retrieval System. These programs and associated linguistic and semantic dictionaries and rule systems are used weekly to process the LP surgical pathology reports for input to a database maintained for the LP within the Clinical Information Utility (CIU). The Pathology Language Encoding System runs on the IBM 370. While queries for the Pathology Retrieval System are generated on the Helix machine, they are executed on the IBM 370.

ISB is currently developing and implementing major enhancements to the Clinical Center, Medical Record Department's Work-in-Process Tracking System to automate additional processes. The functionality of the system is being expanded in three major areas: the database has been redesigned to include additional data elements; processing and accountability of autopsy-related data has been automated and merged with the tracking process; additional reports have been added. The above changes were made in anticipation of the upcoming audit of the Clinical Center by the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO).

ISB staff continue to collaborate with other branches in the DCRT to evaluate technological advances in computer software. During the year, evaluations included: Windows® NT Advanced Server; remote control and remote node software; new Windows® spreadsheets from Lotus® (Improv and 1-2-3

Release 4); and various gateway products to DB2. ISB staff also participated in a formal beta test of a new version of FTP's software.

During FY94 the ISB Applied Systems Programming Section continued to provide support, analysis, design and maintenance services for the following new and longstanding projects:

- the *Visiting International Scientist in America Management Information System* for the Fogarty International Center
- the *Full Time Equivalency Management System* for NIMH, NIAID and OD
- the *NIH Consultant File,* which is used to assist NIH staff in identifying potential members of NIH advisory committees for DRG
- the *Human Leukocytes Antigens Donor System* for the Department of Transfusion Medicine, CC
- continuing development of a database tracking system to record maintenance on all NIH vehicles for the *NIH Transportation Branch*
- the *Human Nutrition Research Information Management System* for the Division of Nutrition Research
- the *Freedom of Information System* that tracks FOI requests to assure that response to these requests are made in a timely fashion, for the OD's Freedom of Information Office
- the *NIH Calendar of Events System* for the OD's Office of Communications
- the *NIH Yellow Pages System* for the OD's Procurement Analysis Branch.

Future Plans

Future plans to improve functionality within the ADB include:

- the incorporation of an Electronic Data Interchange (EDI) between the ADB Procurement system, through the Department's value-added network (VAN), to the vendor
- the development of a system to track small animal purchases by the Veterinary Resources Program of NCRR
- the development of a system for work request entry, estimating, funds approval, billing, tracking of various expenses and reporting for the BEIP within the NCRR
- the development of a system to integrate the usage of credit cards with the ADB Procurement Receiving and Accounts payable functions
- the development of a system to process payments of Research Contracts through the ADB
- the development of an automated process to support the distribution and billing of meal tickets for Clinical Center patients.

The ISB staff will continue to explore how the client/server paradigm can be used to support the strategic and business goals of the NIH. Efforts will focus on cross-platform access to mission-critical data. NSB staff will establish prototype and pilot applications that demonstrate the implementation and utility of client/server architecture at the NIH. To improve the user interface to both the ADB and the ADBIS, we are planning a two- to three-year effort to introduce client/server and Graphical User Interface (GUI) technologies. This effort will improve the user interface for online transaction processing (OLTP) and will provide decision support facilities at the desktop. ISB is currently negotiating contract support to develop these user interfaces.

Given the scope and complexity of the client/server environment, the selection of client tools, connectivity technologies, configuration management technologies, and database servers for evaluation and support will be a collaborative effort among several branches within the DCRT.

ISB staff will continue to offer advice and guidance concerning ICD-specific software development projects, providing expertise in the areas of needs assessment, requirements specifications, database design and development. While tracking and evaluating rapidly evolving database technologies, the ISB will assist with the migration to client/server computing by providing support for various client/server application development tools, as well as query and reporting tools that enable cross-platform access to NIH data. To supplement current services and to further assist the NIH community with their information processing needs, the ISB is currently investigating a contract mechanism to offer professional service referrals for the creation of new database management systems and re-engineering and/or support of existing systems. ISB hopes to supplement its current database and software development services by offering contractor support services in the future.

Closely related to ISB's overall strategic plan is the goal of re-engineering legacy (i.e., existing) systems, in particular, the Administrative Data Base and the Central Accounting System, using systems modeling techniques. In re-engineering, the challenge is to capture the functionality of the current system through a reverse engineering process and then forward engineer the system, applying new technologies. Through this approach, ISB hopes to leverage the investment that has been made in these systems over the years. ISB will

then be in a better position to work on the migration of these systems to new technologies.

As information systems professionals, the ISB staff is further developing its skills and expertise in structured systems analysis and design. In so doing, the ISB staff will be better able to serve the NIH in helping end-users define their requirements. The ISB staff will also be providing support and consultation for the end-users in the planning and analysis phases of the system development life-cycle.

OFFICE OF THE DIRECTOR

Office of the Director

David Rodbard, M.D., Director
William Risso, Deputy Director

The Office of the Director (OD) provides overall program and management direction for DCRT. The Director, Deputy Director, Associate DIrectors, Assistant Directors, and Executive Officer work together as the immediate Office of the Director, whose activities in 1994 encompassed such issues as:

- management of the Division, including allocation of budget, personnel and other resources
- liaison with NIH/OD and all ICDs
- development of new initiatives
- integration of the activities of the offices, laboratories and branches of the Division
- development of necessary data to design and acquire the next generation of technology
- interface with regulatory issues and agencies
- support and guidance for computational molecular biology
- division reorganization
- the High Performance Computing and Communication initiative
- information resources management and strategic planning
- liaison with other Federal agencies.

OD oversees two scientific groups, the *Computational Molecular Biology Section* (CMBS), and the *Unit on Electrocardiography and Signal Processing*. Because of the scientific nature of their contributions to the Annual Report, these are included with those of the OCB, to which they are more closely related than to the administrative activities of the OD.

Three offices supplement the work of the DCRT laboratories and branches:

- The *Equal Employment Opportunity Office* (EEO) manages a full EEO program for the Division. The office serves as the focal point and advisory for all activities relating to the equal employment opportunities of DCRT employees and applicants. The EEO Officer maintains a close working relationship with the NIH Division of Equal Opportunity and other components concerned with minority and women's issues.
- The *Office of Information Resources Managements* (OIRM) is responsible for coordinating and preparing the DCRT contribution to the NIH IRM Strategic Plan, Tactical Plan and the Environment and Resources Report. The DCRT OIRM will focus on major DCRT procurements, providing planning, oversight and technical guidance. ADP security will continue to evolve and require additional attention.
- The *Office of Administrative Management* (OAM) provides administrative and managerial support for the work of DCRT. OAM includes Administrative, Personnel, Financial Management, Project Control and Information Offices, and the DCRT Library.

Equal Employment Opportunity Office

Gloria I. Richardson, EEO Officer

The Equal Employment Opportunity (EEO) Office manages a full EEO program for the Division. The office serves as the focal point and advisory body for all activities relating to the equal employment opportunities of DCRT employees and applicants. The EEO Officer maintains a close working relationship with the NIH Office of Equal Opportunity and other components concerned with affirmative employment issues.

EEO Employee Advisory Committee (EAC)

In its continuing role of providing an important channel of communication and liaison between the director and all division employees, the DCRT EEO Employee Advisory Committee (EAC) (co-chaired by: Susan Chaffee and Delores Stroud) sponsored semiannual Get-Togethers. These sessions are used to increase communication at each level in the Division. Going beyond their call of duty, the EAC helped raise over $700 for Camp Friendship, a camp in Olney, Maryland, for children with cancer.

A major undertaking for the Committee this year was the creation and distribution of an employee questionnaire that dealt with morale, the reorganization, and other pertinent issues. In lieu of a Town Meeting, the EAC met with the Division Director in a Round Table discussion and reviewed the findings gathered from the questionnaire. A summary of the information shared was distributed to DCRT staff through its new listserv list: DCRTStaf. This information was later shared in discussion by the EAC s co-chairpersons at a Lab/Branch Chiefs meeting. As a result, all DCRT staff will continue to work on ways to improve communications within the Division.

As sponsors of the D.C. Partners in Education Program with Wilson High School, the EAC has

continued to support the instructional program at Wilson in the following ways:

- provided technical assistance in planning school-wide computer network
- worked with Wilson to connect them to NIH's EDNET
- provided tours of DCRT facilities to Wilson staff and students
- invited teachers to participate in DCRT training sessions
- furnished computer equipment on loan.

Additionally, the Division eagerly looks forward to serving as judges at Wilson's annual Science Fair. This year, the following persons served: John Burke, James DeLeo, George Dunham, Veronica Nisbet, and Edward Persky.

The EAC sponsored the 30th Anniversary T' Shirt Design Contest, which fielded 28 tee-shirt design entries. There were approximately 230 teal and ash colored T shirts fashioned at this year's picnic!

NIH Target Access Program (TAP)

NIH employees with disabilities or special needs will now find it easier to access the electronic tools that enhance productivity, thanks to a newly signed agreement between DCRT and the U.S. Department of Agriculture. The pact allows NIHers to use USDA's Target Center, a state-of-the-art demonstration facility for enabling technologies.

The NIH Target Access Program (TAP) was developed by Networks Systems Branch Computer Specialist, Bronna Cohen, who saw a need for such a program.

The Target Center is located at the USDA Headquarters South Building in Washington, D.C., immediately adjacent to the Smithsonian Metro station. The Center is fully accessible. To make an appointment, or for further information, please call 301-594-DCRT (TTY 301-496- 8294), or send electronic mail to 4DCRT@nih.gov.

Minority Programs Symposium

DCRT participated in the 1993 Minority Programs Symposium, sponsored by the National Institute of General Medical Sciences. The symposium was hosted by Morehouse School of Medicine at the Atlanta Hilton and Towers in Georgia. Richard J. Feldmann, Computer Specialist, served as a panelist on the Computational Sciences/Imaging Workshop. Helen Madigan-Sedor, Personnel Management Specialist, performed public relations work in the exhibit area by sharing brochures about the organization; by emphasizing how the DCRT collaborates in molecular biology, chemistry, physics, mathematics, and medical imaging; receiving applications for summer employment; and by talking to students regarding career opportunities in the Division as well as throughout the NIH.

EEO Award Recipient

The Division's 1993 EEO Award recipient was Reneé Edwards. Ms. Edwards is a Computer Specialist with the Information Systems Branch. The award was given in appreciation of her creativity, insight, and good judgment demonstrated in addressing EEO concerns and for the significant effort, time, and talent contributed during her tenure as the Division's EEO Representative to the NIH Black Employees Advisory Committee and to the DCRT EEO Employee Advisory Committee.

The 1994 Director's EEO Special Achievement Awards were granted to one non-supervisory employee and to two managers. The recipients were: Bronna Cohen, computer specialist, NSB, for her initiation, coordination, and implementation of TAP; Kevin Murphy, Personnel Officer, for his special role in the design of the DCRT Career Enhancement program (CEP), which gives employees in technical and administrative positions the opportunity to move into a professional series with greater career and promotion potential; and Robert Mamayek, Chief, Systems Operations Management Section, CFB, for his implementation of the CEP.

Office of Information Resources Management

Arthur Schultz, Chief

The DCRT Office of Information Resources Management (OIRM) supports the information technology needs of the NIH intramural, extramural, and administrative communities. The Division has accepted a central role in the IRM process at NIH by developing a strategic architectural plan for information technology. In conjunction with the ICDs, DCRT has prepared a Technical Architecture Plan, with the goal of standardizing many common administrative processes across the NIH for increased interoperability. DCRT also will work toward helping NIH achieve a corporate view of computing architectures and standards, and will continue to advocate and negotiate for adequate resources to build and maintain the strong computational presence required for modern biomedical research. DCRT will

encourage purchases that are designed for operation in a heterogenous environment. OIRM has coordinated the preparation of IRM submissions by DCRT to the overall NIH Strategic Plan. These submissions included the IRM Strategic Plan, 1996-2000, the IRM Tactical Plan which includes the Information Technology Budget, and the Environment and Resources Report.

CERTAN Acquisition

The "Computer Equipment, Resources and Technology Acquisition for NIH" (CERTAN) will provide scientific and administrative computer systems for which DCRT provides support beginning in 1996. DCRT requested the General Services Administration (GSA), Federal Systems Integration and Management Center (FEDSIM), to assist in: (1) developing a user-based Concept of Operations (CONOP); (2) conducting a requirements analysis; and (3) preparing a Statement of Work (SOW) for the contracts. The user-based CONOP, completed in May 1993, made recommendations for improving the current system, and identified areas in which users want DCRT to provide central computing support.

Sixteen workshops and 32 interviews were conducted with a broad spectrum of NIH clinicians, scientists, administrators, and managers, totaling 327 participants. These meetings were conducted to expand the requirements, focusing on the computing support areas identified in the CONOP. The user-input collected during the workshops and interviews form the basis for the requirements analysis. The workshops and interviews were facilitated by a FEDSIM contractor, who prepared a written summary, which was reviewed and edited by the participants. The acquisition team recommended NIH organizational units with specific responsibilities in the areas identified by FEDSIM for DCRT to provide central computing support.

The DCRT OIRM will have a unique role in the CERTAN acquisition. As the planning coordinator for DCRT, the OIRM will concurrently address Division-wide IRM business and technical issues, while encouraging Institutes, Centers, and Divisions (ICDs) to participate in the planning process.

The requirements analysis is divided into four focus areas, matching the anticipated four contracts. As a member of the CERTAN acquisition staff, OIRM's Chief serves as chairman of the group reviewing the requirements analysis for the Support Services contract, and will also be one of the two deputy trail bosses focusing on technical issues. He completed the "Trail Boss" Training Seminar for Managers in December, 1993.

It is envisaged that the scope of the Support Services contract is to provide the NIH ICDs with a vehicle to acquire needed services on a task order basis. In addition, it will provide resources for the entire Division, commensurate with and, in many cases, exceeding services provided by a variety of existing smaller contracts. The CERTAN acquisition will provide services that currently are furnished by a variety of existing contracts and in-progress acquisitions. As a part of the overall division planning, contracts that will be supplanted in the future by CERTAN have been identified, and the CERTAN acquisition has been reduced by the value of these contracts until they expire. One technique DCRT has employed to encourage interest in the CERTAN acquisitions is to schedule one-hour discussions with vendors who have responded to the Commerce and Business Daily notice announcing CERTAN. A maximum of four vendors are scheduled each Friday afternoon. The vendors meet with a technical member and an IRM member of the acquisition staff. The vendors are given the opportunity to present the capabilities of their company and to ask questions concerning the progress of the acquisition.

OIRM Staff

A computer specialist was recruited from within the Division to augment the OIRM staff. She will be responsible for coordinating the software licensing initiatives within the Division and will become a member of the committee reviewing the requirements and developing a "Section C" for the CERTAN Support Services Contract.

Software License

An NIH-wide policy for approving software license agreements was established last year. This process requires that the license be signed by the contracting official, and was based on a recommendation by the Office of the General Counsel that the NIH contracting official responsible for the software acquisition review and sign licenses. At this time the process is extremely slow, in part because the license content must be negotiated between the vendor and the contracting official and further between the General Counsel and the contracting official. The prospect exists for more timely renewals after the initial license, based on the vendor and the contracting official agreeing to accept the previous license.

Computer Emergency Response Team

OIRM oversees and coordinates DCRT initiatives and cross-platform issues regarding computer security. Beginning in FY94, a DCRT employee was detailed to provide technical and administrative support to the DCRT OIRM on security issues and has provided the leadership to establish a DCRT Computer Emergency Response Team (CERT). The CERT has provided coordination and communication among DCRT staff during intrusion incidents involving DCRT computing systems. The CERT coordinated the response to the discovery in March, 1994, of a "sniffer" program installed on an NIH platform, and has worked with the NIH OIRM and various law enforcement agencies in the investigation of other incidents.

The DCRT OIRM initiated a security review of DCRT by the Carnegie Mellon CERT (a federally-funded group responsible for overall Internet security). The Carnegie Mellon CERT review will provide recommendations for improving the security environment of DCRT systems and techniques for offering better security for resources and services to the NIH campus. In addition, during the coming year, DCRT will investigate commercial security consulting services with the objective of contracting with a security consultant for advice in some areas of computer security.

DCRT OIRM collaborated with the NIH OIRM to develop a network risk assessment methodology for NIH LANs. This unique self-assessment methodology will enable LAN administrators to assess the security levels of their PC-based LANs. OIRM is evaluating the possible need for an automated version of the risk assessment methodology.

TARGET ("Technology Accessible Resource Gives Employment Today") Access Program

DCRT has for some time recognized the need to provide leadership in making computer-based assistive technologies available to NIH employees with disabilities. This need became compelling with the signing of the Americans with Disabilities Act in 1990. DCRT signed an interagency agreement with the Department of Agriculture (DOA), and inaugurated its sponsorship of the NIH TARGET Access Program (TAP). TAP is intended to assist NIH employees with disabilities or special needs find the tools they need to be productive.

TAP allows NIH employees to share the DOA TARGET Center, a state-of-the-art demonstration facility, which features technologies designed for persons with disabilities. TARGET Center users can evaluate the best available technology. The Center's full-time staff consults with visitors and callers to assist them in finding accommodations. TAP has received a warm welcome from NIH employees and enthusiastic support from the NIH OEO and EEO community. The NIH Division of Safety plans to utilize the program to provide specialized supplemental training for its occupational safety and health consulting staff. In addition, the NIH Occupational Medical Service plans to use TAP as a referral resource for employees who experience injury or disability. This project has been a joint effort of DCRT OIRM, the Employee Advisory Committee (EAC), OEO, CSB and NSB.

Office of Administrative Management

Marian L. Dawson, Chief

The Office of Administrative Management (OAM) provides guidance and support on all administrative and business management aspects of the Division's programs, advising on the management of resources, the provision of administrative services, program planning and evaluation, and policy and legislative analysis. In addition, the OAM organization includes the DCRT Information Office and the DCRT Library.

Administrative Management Section

The Administrative Management Section (AMS), with a staff of seven, headed by Administrative Officer Marlyn Harrison, continues to provide administrative services to the Division, with particular emphasis on contracting and procurement, property reconciliation and inventory control, timekeeping, training, and space, with other administrative functions necessary for the efficient and effective implementation of program objectives.

Last year, the AMS worked with the Information Systems Branch to bring the DCRT branches on-line with the new NIH Administrative Data Base feature, Request for Purchase Action (RPA), for procurement activities. This area continues to evolve, with the Division providing labs/branches with a more efficient method of procuring supplies and services. This year DCRT underwent reconciliation audits from the Public Health Service and the NIH Chief Financial Officer

(CFO) regarding property management. The success of these audits placed DCRT below the 2% level requirement for lost or unaccountable property. Also, the AMS coordinated an effort to bring DCRT on-line with timekeeping (through TAIMS) and training (through IMPACT, now in its pilot stage). The AMS has worked extensively with the Division of Engineering services (DES) to facilitate the construction/conversion of storage space in the 12B basement to offices and a conference facility. In addition, renovation of the 12A conference room has been planned, with completion of both projects anticipated for early FY95.

DCRT Budget Office

The Budget Office, headed by Mr. Michael Reed, continued to carry out its financial functions for the Division, including budget formulation for its two funding mechanisms, and preparation for budget review by the NIH Central Services Budget Review Committee. In cooperation with the DCRT/OIRM, the Budget Office supplied budgetary information for the Information Technology Systems (ITS) budget, submitted through NIH/OIRM to PHS and DHHS. The office also administered the allocation of available funds within the Division, tracked expenditures for each organizational area, and provided various financial reports, reviews and analyses for DCRT management. A system for providing budget execution reports to program managers was developed, using a combination of Central Accounting System and Administrative Database Information System (ADBIS) data. The office also coordinated financial aspects of proposed cost recovery plans for the Helix project and for Networking. It also responded to requirements for Internal Control Reviews and supplied materials to NIH/Division of Financial Management for the annual CFO audit.

Human Resources Management Section

The DCRT Human Resources Management Section (HRMS), headed by Kevin Murphy, advises and assists management in providing and optimally utilizing human resources in order to accomplish Divisional goals, and is responsible for conducting the Human Resources Management Program for the Division. This includes staffing and recruitment services, compensation and classification, employee benefits programs, supervisory/employee development, management analysis, performance appraisal, awards and incentive programs, employee relations, conduct and ethics issues.

This year, the office was again influential in several key initiatives, including the reorganization of the entire Division. The Human Resources Management Section was an integral player in the intensive review of DCRT programs, and wrote position descriptions, numerous internal vacancy announcements for new positions and determined position management strategies and staffing patterns for the new organizations. Additionally, we changed our name (previously the Personnel Operations Section) to reflect our goal of providing DCRT's employees and supervisors with the full range of human resources needs and services.

The DCRT Career Enhancement Program (CEP) is an upward mobility program designed by this office and geared towards creating professional opportunities for technical staff. In the second year of the program, five employees were selected for computer specialist positions within the Division. The HRMS staff wrote the formal training plans for the participants and coordinated with the supervisors to ensure effective implementation of the program.

The NIH hiring freeze is a direct result of the federal government's effort to streamline agencies and programs. However, the HRMS was able to augment Division staff by using exempt programs. DCRT continues its affiliation with the "Partners in Education Program" at Woodrow Wilson High School, Washington, DC. The partnership agreement formalizes DCRT's commitment to attracting and recruiting talented young people, and to furthering science education at the secondary school level.

The HRMS provided an increasing variety of assistance and information on employee benefits and services, including the open-seasons for federal health insurance and the Thrift Savings Plan. The staff also greatly facilitated the expansion of the Alternative Work Schedule/Work Place Program throughout the Division. HRMS specialists also wrote and processed many honor and cash awards for outstanding employees and work groups. The Management Analyst wrote and disseminated over five administrative policy and procedure statements for the Division, including leave administration, outside activities, and a group training policy.

The NIH, including DCRT, was granted by Congress in May authority to offer early retirement and buyout incentives to eligible employees. The HRMS staff determined that over 50 DCRT employees were eligible for the buyout incentive. The annuity estimates were then calculated for each of these

employees and retirement consultations were conducted. In all, DCRT reduced its FTE count by four retirees. The office is preparing for the likelihood that another round of buyouts will be offered in October, 1994.

The HRMS staff was responsible for ensuring that all of our approximately 320 employees attend the new mandatory AIDS/HIV Awareness Training. The Management Analyst coordinated two sessions of group training for excellent customer service skills and an in-house session on time and attendance training for leave approving officials.

The HRMS is preparing for the next round of changes in quality management and organizational structure. We anticipate more opportunities to redelegate HRM related functions to supervisors and managers. For example, we are readily available for guidance on developing the newly required Career Development Plans for all employees. These plans are designed to open communication between the employee and the supervisor regarding needs as well as personal and professional goals.

Lastly, the HRMS staff has attended specialized training courses, seminars and professional conferences in order to sharpen skills and keep abreast of the many new developments in Human Resources. These include the American Society for Public Administration Conference on "The Managerial Presidency" in March 1994, the U.S. Office of Personnel Management Symposium on Labor and Employee Relations in January 1994, and the International Personnel Management Association Eastern Region Conference in Rockville, Maryland in June 1994.

DCRT Library

Ellen Moy Chu, Chief

The DCRT Library provides information resources to support NIH research and development in the subject areas of computer science, mathematics, and computer applications. Information technologies are transforming libraries from print repositories to virtual libraries, providing information in digital formats locally, or to users' desktop computers at remote sites. The library staff of two librarians and two part-time students continues to evaluate, implement, and refine electronic information delivery systems for enhanced user services.

In FY94, virtual library activities introduced new information delivery platforms and programs:

- *STILAS*: The library integrated system application, Scientific and Technical Information Library Automation System (STILAS) now provides online catalog, circulation, and serials control. This year, staff implemented reporting features in these modules and initialized the acquisitions module to migrate from a dBASE system. This brings together order and inventory information for library books, journals, microforms, and digital publications in a centralized database. NIH staff and Internet users daily access the online catalog available 24 hours on the NIH GOPHER.

Two new modules allow users to search other databases with the online catalog and to access remote databases via a multiuser gateway. Current Index to Statistics (CIS/ED) is the first local database. End users now search for statistical bibliographic citations, and then select the online catalog database to determine availability in the DCRT Library collection. NIH users have a special login to access this version of the STILAS catalog on Internet or on the NIH GOPHER. The multiuser gateway will provide end users with Internet and/or dial access to the Online Computer Library Center (OCLC) in Ohio or to Dialog in California. These bibliographic utilities provide access to hundreds of databases, which contain full text and/or bibliographic citations to research and technology subjects. The library will profile access to these systems to meet user interests and needs.

STILAS operates server functions on a SUN® SPARC2. A year ago, the Computer Facilities Branch (CFB) staff began testing STILAS in an Advanced Laboratory Workstation (ALW) configuration to investigate the feasibility of operating STILAS in the ALW environment. Testing by ALW staff will resume when time permits. A successful migration to the ALW environment will provide system maintenance, security, and software upgrades. After the NIH May security alert, ALW and Distributed Systems Branch (DSB) staff inspected the workstation and installed a program to provide an automated weekly report to monitor system files.

Staff downsizing led to purchase of remote journal check-in service. The contractor in Virginia accesses STILAS on the Internet, enters check-in information into the database, and prints out routing slips. Journal issues are prepared for public use and shipped to the Library. This has provided timely information delivery in this vital scientific publication format.

- *Table of Contents E-mail Pilot Project*: Computational Bioscience and Engineering Laboratory (CBEL) staff selected journal titles in this new OCLC ContentsAlert service. E-mail on Internet delivers copies of tables of contents from Ohio directly to individual participants. This

alerting service enables CBEL staff to keep up with rapid changes in their research areas without visiting the library to check journal issues, and provides information from journals not owned by the library. A review and evaluation of customer satisfaction will determine if this service will be extended to other laboratories and branches.

- *CD ROMs*: Campus-wide access to our most popular CD ROM™ publication, Computer Select, changed to accommodate migration to the LAN Manager Network Operating System and Windows®. The Library continues to cooperate with DSB staff operating PUBnet, which provides CD ROM™ access on NIHnet. Computer Select provides full-text articles from personal computer (PC) and Macintosh® journals, and directories of software and hardware products and vendors. Within the library, users may access Books in Print Plus and CRISP on a public workstation.
- *Public Workstations*: Upgrades of both the Macintosh® and PC workstations for public use enabled or enhanced user access to Internet navigation tools, such as MOSAIC and GOPHER. Users may also search MEDLINE using GratefulMED on either platform. The PC provides access to the CD ROM publications on PUBnet.

Training programs were a major focus of FY94 library activities. On October 15, an Open House to demonstrate operation of the new compact/movable shelving provided hands-on training to library users. An article in the NIH Record generated calls from interested NIH offices, and the NIH Library inspected the system twice to prepare their procurement.

DCRT participates with Woodrow Wilson High School in a Partners in Education program. In November, both high school librarians visited the DCRT Library to determine what information resources might be beneficial for their students. DCRT librarians then visited Woodrow Wilson's library and computer laboratory. We loaned a PC and printer to the Woodrow Wilson library for their information applications. DCRT staff regularly donate journal issues and computer catalogs to the high school library.

Orientation programs began with an introduction to library services for new NIH staff in the October Welcome to DCRT program. As part of the DCRT training program, a hands-on class in March provided students with opportunities to search the online catalog, the Computer Select CD ROM, and GratefulMED. The August class offered a tour of the library and its facilities.

Library staff communicated service and acquisition information to users with articles in PCBriefs and within the Information Desk module of STILAS. The Chief submitted annual reports to the DCRT Office of Information Resources Management and the NIH Privacy Act Coordinator. Since January 1994, over forty consultants and company representatives have registered to use the reading file of DCRT background information for the Computer Equipment Resources and Technology Acquisitions (CERTAN).

The Chief provided consultative services to the Division of Engineering Services staff contracted to set up a library within DES. As the DCRT representative to the NIH Library Advisory Committee, the Chief works with various DCRT components to facilitate collaborations with the NIH Library in providing NIH-wide information services. NIH Library staff, CFB GOPHER/MOSAIC staff, and the DCRT Library Chief meet bimonthly to coordinate activities. Advance notice of plans coordinates procurement to avoid duplication, and also enables NIH Library staff to anticipate user reference queries.

Both librarians were engaged in DCRT Employee Advisory Committee activities this year. A major project was the development, analysis, and communication of results of the employee survey. Staff continue to represent DCRT at the NIH Advisory Committee for Women. This year, the librarians attended training courses in supervision, management, customer service, retirement, ethics, and HIV/AIDS, and in computer areas such as UNIX®, computer security on the Internet and related to the Privacy Act, GOPHER/MOSAIC, WINDOWS®, WordPerfect®, and CURE, BRMUG, and PC Topic Sessions. Library specific training included: STILAS acquisitions, serials, and User Group meetings, MEDLINE, GratefulMED for PC and Macintosh®, Internet library services, and site visits to libraries at the Justice Department, Pentagon, and the Naval Research Laboratory.

Next year, we look forward to developmental work with multimedia electronic publishing and enhanced campus-wide information systems, as resources permit.

DCRT Information Office

The DCRT Information Office (IO) is responsible for communicating the advantages of computer research and technology to NIH administrative and scientific communities; electronic and print media, including the

trade press; users of DCRT services; and the general public. Equally important is the responsibility for helping to build and maintain communication within DCRT. To accomplish these objectives, a staff of two public affairs specialists are involved in the following functions:

- writing and editing
- graphic arts and photography
- publications
- special events
- support for the Office of the Director and the larger NIH community
- internal communications
- media liaison
- public inquiries.

DCRT's 30th Anniversary

The year 1994 marked the 30th anniversary of the founding of DCRT, and the office was heavily involved in commemorative activities. An extensive article on DCRT's past, present and future was produced for the NIH Record and reprinted for additional uses. On May 2, 1994, a well-attended DCRT 30th anniversary symposium was held in Masur Auditorium, with Drs. Russell Doolittle of the University of California, San Diego and Martin Karplus of Harvard University making major presentations. Dr. Rodbard briefly reviewed the history, accomplishments and contributions of DCRT. In June, a special version of the annual DCRT picnic was held, featuring Division alumni, awards and tee-shirts, a birthday cake, and games and activities with a "30" flavor.

Writing, Editing, and Publications

The office produced an indexed, desktop version of the *1993 Director's Report* and a draft of the 1994 report, and completed an update of the *NIH Directory of Image Processing Facilities*, and began revising DCRT's *Computing Resources*. A major feature story was produced on contract for the NIH Catalyst about the Computational Bioscience and Engineering Laboratory; additional feature stories are planned on computational molecular biology, DCRT and employees with disabilities, and Orphanet, a network linking individual scientists not connected to a local area network. Submissions to the NIH Record from DCRT staff included an article on such subjects as the new e-mail directory, a "sniffer" attack on the Internet; new directions for the Scientific Computing Resource Center, the TARGET CENTER for employees with disabilities; and numerous shorter stories and photo/captions related to Division training, awards, and other activities. In addition, the Information Office contributed updates to the Biennial Report, Scientific Directory/Annual Bibliography, international and audiovisual reports, the NIH Almanac, the NIH Calendar of Events and Meetings, and the Postdoctoral Research Fellowship Opportunity Booklet.

Graphic Arts and Photography

Fiscal Year 1994 saw the Information Office remain heavily involved in graphic arts and photography. Numerous photo sessions were conducted, including an all-DCRT portrait to mark the 30th anniversary. Posters, tent cards, fliers, and programs were produced in conjunction with NIH's Medical Arts and Photography Branch for the 30th Anniversary Symposium. Special displays produced by the office included:

- the Computational Bioscience and Engineering Laboratory
- computational molecular biology
- DCRT Director's NIH Award winners
- DHHS Distinguished Service Award
- the DCRT 30th Anniversary Symposium
- several DCRT scientists and their prominent papers.

The office also added more large flannel boards in Bldg. 12A to display posters from the NIH Research Festival.

Special Events

Staff members organized the highly successful 30th Anniversary Symposium, Division picnic, and "Welcome to DCRT" presentation for new postdoctoral fellows; and assisted with the annual DCRT Awards Ceremony. The office was also called upon to help conduct visits by Wilson High School students and guests from the People's Republic of China and from Taiwan.

OD Initiatives

In addition to providing background information, slides, audio/videotapes, publication packets, and other materials for OD, the office increased campus distribution of DCRT publications and encouraged submission of more award nominations. Information Office staff also worked on NIH committees to increase the availability of candidates for minority public affairs specialists and to develop electronic means of disseminating patient education materials.

Media Activity

Information about DCRT was provided to the following media organizations:

- Science, a journal of the American Association for the Advancement of Science
- Federal Computer Week
- The Wall Street Journal
- Public Broadcasting System
- Washington FAX
- Modern Healthcare Magazine
- Science and Technology Satellite News
- Electronic Health Records Report
- U.S. Pharmacist Magazine.

Acknowledgments

The Director, DCRT, wishes to acknowledge the valuable contributions of J. Emmett Ward, V. Adrian Parsegian and Mary W. Hodges in the preparation of this publication.

Acronyms

Å	Ångstrom
ABI	Applied Biosystems, Inc.
ACH	Automated Clearing House
ACR-NEMA	American College of Radiology-National Electrical Manufacturers Association
ACTH	adrenocortioctropic hormone
ADAMHA	Alcohol, Drug Abuse, and Mental Health Administration
ADB	Administrative Data Base
ADBIS	Administrative Data Base Information System
ADP	Automatic Data Processing
AECG	Ambulatory Electrocardiography
AFS	Andrew File System
AIMS	Animal Information Management System
ALW	Advanced Laboratory Workstations
AmB	Amphotericin B
AMS	Administrative Management Section
ASA	accessible surface area
ATM	asynchronous transfer mode
BDAM	Basic Direct Access Storage Method
BEIP	Biomedical Engineering and Instrumentation Program
BIMAS	BioInformatics and Molecular Analysis Section
BOSS	Best of Open Systems Solutions
BPB	Biological Psychiatry Branch
BPMS	Brain Position Measurement System
bpv	bovine papilloma virus
BRMUG	Biomedical Research Macintosh Users Group
BRB	Bone Research Branch
CAP	Cluster Analysis Program
CAS	Central Accounting System
CBEL	Computational Bioscience and Engineering Laboratory
CC	Clinical Center
CD-ROM	Compact Disk-Read Only Memory
CERT	Computer Energency Response Team
CERTAN	Computer Equipment, Resources and Technology Acquisition for NIH
CFB	Computing Facilities Branch
CFO	Chief Financial Officer
CHIPS	Child Health Information System
CIU	Clinical Information Utility
CMBS	Computational Molecular Biology Section
CMS	Capacity Management Staff
CONOP	Concept of Operations
CPU	Central Processing Unit
CRADA	Cooperative Research and Development Agreement
CRISP	Computer Retrieval of Information on Scientific Projects
CRH	corticotropin-releasing hormone
CSAR	Computational Sequence Analysis Resource
CSB	Customer Services Branch
CT	Computed Tomography
CURE	Campus Users Resource Exchange
DASD	Direct Access Storage Device
DASMOC	dihedral angle space Monte Carlo
DCBDC	Division of Cancer Biology, Diagnosis, and Centers
DCE	Distributed Computing Environment

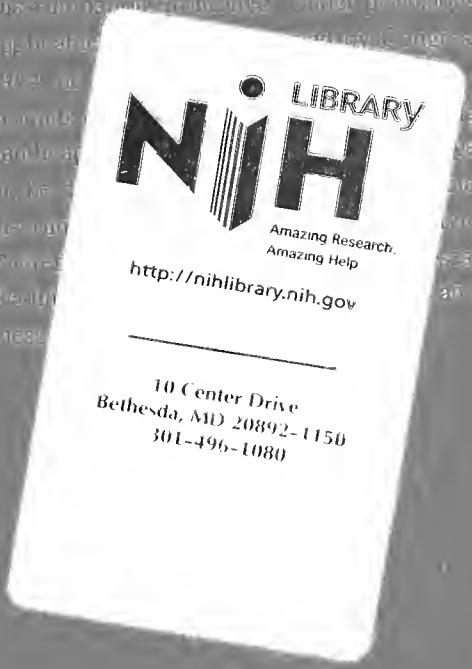
DCPC	Division of Cancer Prevention and Control
DCRT	Division of Computer Research and Technology
DCT	Division of Cancer Treatment
DFM	Division of Financial Management
DFT	Discrete Fourier Transform
DMS	Data Management System
DNA	Deoxyribonucleic Acid
DNM	Department of Nuclear Medicine
DOS	Disk Operating System
DPM	Department of Personnel Management
DRD	Diagnostic Radiology Department
DSA	Digital Subtraction Angiography
DSB	Distributed Systems Branch
DSS	Distributed Systems Section
EAC	Employee Advisory Committee
ECG	Electrocardiogram
EEG	Electroencephalogram
EEO	Equal Employment Opportunity
EM	Expectation Maximization
EPA	Environmental Protection Agency
EPS	Enterprise Print Services
ERI	electron repulsion integral
FCCSET	Federal Coordinating Council for Science, Engineering and Technology
FDDI	Fiber Distributed Data Interface
FEDCAC	Federal Computer Acquisition Center
FEDSIM	Federal System Integration and Management Center
FFT	Fast Fourier Transform
FNS	Fiber Network Services
FOV	Field of View
FTEs	Full Time Equivalents
FTP	File Transfer Protocol
FY	financial year
GCG	Genetics Computer Group
GSA	General Services Administration
GUI	Graphical User Interface
HFSS	Hierarchical File Storage System
HPA	hypothalamic-pituitary-adrenal
HPSCS	High Performance Scientific Computing Section
HSM	Hierarchical Storage Manager
ICDs	Institutes, Divisions and Centers
IMPAC	Information for Management, Planning, Analysis, and Coordination
I/O	Input/Output
IPRS	Image Processing Research Section
IRM	Information Resources Management
IRP	Intramural Research Program
ISB	Information Systems Branch
ITC	Image Technology Center
LAN	Local Area Network
LAP	Laboratory Analysis Package
LBM	Laboratory of Biochemistry and Metabolism
LCAS	Laboratory and Clinical Applications Section
LCB	Laboratory of Chemical Biology
LCB	Laboratory of Cell Biology
LCE	Laboratory of Comparative Ethology

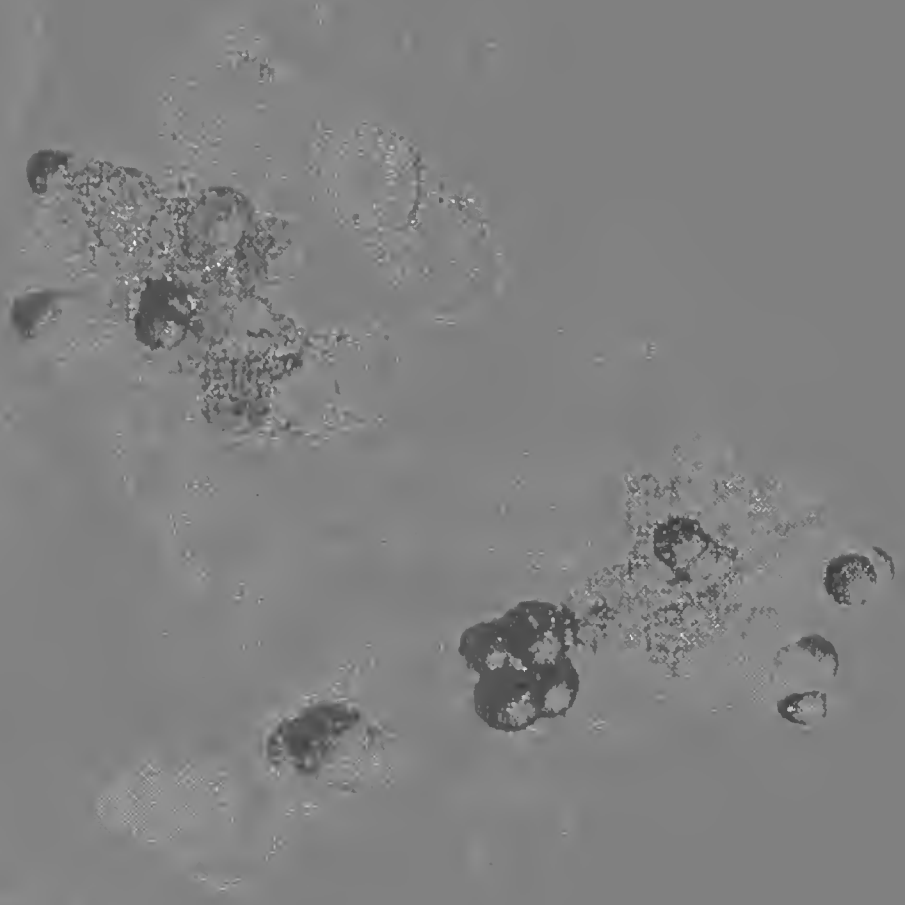
LDACS	Laboratory Data Acquisition and Control System
LDRR	Laboratory of Diagnostic Radiology Research
LMB	Laboratory of Molecular Biology
LSB	Laboratory of Structural Biology
LTIB	Laboratory of Tumor Immunology and Biology
LTPB	Laboratory of Theoretical and Physical Biology
LTS	LAN Technology Section
Mac	Macintosh®
MIMD	Multiple Instruction Stream/Multiple Data Stream
MPP	massively parallel processors
MSS	Managed Storage System
MVS	Multiple Virtual Storage
NASA	National Aeronautics and Space Administration
NCBI	National Center for Biotechnology Information
NCHGR	National Center for Human Genome Research
NCI	National Cancer Institute
NCRR	National Center for Research Resources
NEI	National Eye Institute
NHLBI	National Heart, Lung and Blood Institute
NIA	National Institute on Aging
NIAAA	National Institute on Alcohol Abuse and Alcoholism
NIAMS	National Institute of Arthritis and Musculoskeletal and Skin Diseases
NICHD	National Institute of Child Health and Human Development
NIDA	National Institute on Drug Abuse
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NINDS	National Institute of Neurological Disorders and Stroke
NINR	National Institute for Nursing Research
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
NOS	Network Operating System
NQS	Network Queuing System
NRCC	Naval Regional Contracting Center
NSB	Network Systems Branch
NTAS	Windows NT Advanced Server
OAD	Office of the Associate Director
OAM	Office of Administrative Management
OCB	Office of Computational Biosciences
OCRS	Office of Computing Resources and Services
OGCS	Ophthalmic Genetics and Clinical Services Branch
OEO	Office of Equal Opportunity
OIRM	Office of Information Resources Management
OPM	Office of Personnel Management
OSTP	Office of Science and Technology Policy
OTC	Office Technology Coordinators
PBQS	Parallel Batch Queuing System
PC	personal computer
PCO	Project Control Office
PCR	Polymerase Chain Reaction
PDB	Brookhaven Protein Databank
PEGs	Polyethylene Glycols
PET	Positron Emission Tomography
PIG	Product Information Guide

PMT	Photomultiplier Tube
POP	Post Office Protocol
PPP	Point-to-Point Protocol
PSL	Physical Sciences Laboratory
PUBnet	NIH Public Network
QM	Quantum Mechanical
QM-MM	Quantum Mechanics-Molecular Mechanics
QMF	Query Management Facility
RACF	Resource Allocation Control Facility
RAID	Redundant Arrays of Inexpensive Discs
RAM	Random Access Memory
RCOMM	Remote File Access and Communication System
RCWS	Radiology Consultation Workstation
RFP	Requests For Proposals
rms	root mean square
ROB	Radiation Oncology Branch
RPA	Request for Purchase Action
RPC	Remote Procedure Call
ROC	Receiver Operating Characteristic
RRV	R-R Interval Variability
SAS	Statistical Analysis System
SCF	Self-Consistent Field
SCRC	Scientific Computing Resource Center
SCS	System Consulting Section
SGI	Silicon Graphics Indigo
SMS	System Managed Storage
SOMS	Systems Operations Management Section
SOW	Statement of Work
SQL	Systems Query Language
SPECT	Single Photon Emission Computer Tomography
SPM	Statistical Parametric Mapping
SRM	System Resource Manager
SSFAS	Service and Supply Fund Activity System
SSS	Statistical Support Staff
STILAS	Scientific and Technical Information Library Automation System
SVD	Singular Value Decomposition
TASC	Technical Assistance and Support Center
TCP/IP	Transmission Protocol/Internet Protocol
TLC	Technical LAN Coordinator
UPS	Uninterruptable Power Supply
URC	User Resource Center
VRP	Veterinary Resources Program
VSAM	Virtual Storage Access Method
WAN	Wide Area Network
WTCDC	Washington Theoretical Complex Disease Consortium
WUG	Wordprocessing Users Group
WWW	World Wide Web

Index

*U.S. GOVERNMENT PRINTING OFFICE: 1995-386-931/33580


LIBRARY
NIH
Amazing Research.
Amazing Help
http://nihlibrary.nih.gov
10 Center Drive
Bethesda, MD 20892-1150
301-496-1080



DCRT

Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland 20892